Measurement of Mid-rapidity Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 200$ GeV

Dmitrii Kalinkin

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Physics,
Indiana University
August 2022

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Scott W. Wissink, Ph.D. (Chair)

Emilie Passemar, Ph.D.

Rex Tayloe, Ph.D.

Sabine Lammers, Ph.D.

April 19, 2022

In memory of my dad Igor. I know how much you were looking forward to this.

Acknowledgments

It's been a long way for me to finish this thesis work. I would like to acknowledge and thank some of the people who, through our interactions, inspired and supported me in my life and research pursuit.

Family and friends

I would like to take a moment to appreciate my Mom for her loving care. And I would not have gotten this far without the infinite support and encouragement of my Dad. I thank my dear friends Yahor and Ella for those amazing camping trips and for being someone I can rely on. Special thanks goes to Salvatore for having my back during 2020, for being a good sportsman in archery and all the other nice activities. Thanks to Stephen for always being the most amazing person. I'm sending my love to Priyom for all the great times hanging out on the IU campus. My gratitude for Masha who made Yale University so amazing for me, and for being the soul of the company. Cheers to Tomas for being a kind roommate. Kudos to Sasha for being my accountability partner during the writing of this thesis.

Dear colleagues and collaborators

I would like to thank Dmitri Smirnov for our fruitful conversations. I would like to thank our group members for kind interactions: Akio Ogawa, Ana Nunes, Brian Page, Grant Webb, Huanzhao Liu, Jaroslav Adam, Joseph Kwasizur, Minghui Zhao, Oleg Eyser, Pia Maria Zurita, Salvatore Fazio, Ting Lin, Wan Chang, Wlodek Guryn, William Schmidke, Xiaoxuan Chu and Zhanwen Zhu.

My data analysis relied on the work by Kevin Adkins who performed the energy calibration of the BEMC, data and embedding QA and the initial setup needed for the jet reconstruction. I would like to thank Zilong Chang for all fruitful discussions we had during our research, as well as sharing his work on the collider luminosity, the BEMC response to hadrons, and for his kind support.

Advisors

I would like to thank my dear undergraduate advisors Dmitry S. and Igor A. for checking up on me. I would like to thank Scott Wissink for his wise advising on my research, for his calm and efficient approach to resolving all the matters that occur in the life of a graduate student. I would like to thank Elke Aschenauer for funding my PhD. It was a great honor to

work with Dr. Aschenauer, I've learned a lot of things that I wouldn't have learned from anyone else. I am also very grateful for all the trust and patience she has shown towards me.

Miscellaneous

Thanks to Nelson Batalon for his help with administrative things at IU. I'm thanking Rex Tayloe for being my graduate advisor for all these years. I appreciate every member of my graduate research committee for their kind feedback regarding my research. I would like to thank STAR experts Alexei Lebedev, Gene Van Burren, Jason Webb and Yuri Fisyak. I'd like to thank SDCC staff at BNL and NERSC staff at LBNL (specifically, Jan Balewski and Jeff Porter).

I acknowledge support from the U.S. Department of energy under contract number de-sc0012704.

Table of Contents

Chapter 1

Introduction

1.1 Description of fundamental interactions and structure in nature

Nature is rich in phenomena, and we still have a lot to learn about them. There are several ways in which matter interacts with itself, which produces myriads of effects: from formation of galaxies in space to the chemistry of life on Earth to conductivity of semiconductors in microchips. All of this complexity emerges from only a few fundamental interactions, which in the Standard Model we believe to be:

1. Gravitational interaction
2. Electroweak interaction[1]
3. Strong interaction

The most viable approach to study the interactions themselves is to look at the most simple systems. Many-body systems are known to be hard to solve analytically, so we generally prefer to avoid those when possible. A simpler case study would be two-body systems, for which there would be roughly two kinds of regimes of interaction:

1. *A bound system*: all the interacting bodies occupy a limited region in a physical space (low energy) for a certain, possibly unlimited, amount of time
2. *A scattering system*: initially non-interacting bodies interact briefly, then leave the interaction region to never interact again (high energy)

The initial observations of celestial bodies that are bound to the Sun led to the discovery of Newton's inverse-square law of gravitation. Later, Coulomb's electrostatic forces were first quantified using an experimental setup with two electrically charged spheres and a torsion pendulum in a cage. From a physics standpoint, it was not the most simple system, but it still allowed us to gain insight into the fundamental electromagnetic force. That insight

[1] Which combines both electromagnetic and weak interactions.

was one of the keys to interpreting the famous Geiger-Marsden scattering experiment[2] that allowed us to pinpoint the structure of atoms.

From the philosophical standpoint, what allowed us to gain insight into the workings of fundamental interactions was understanding the structure of the interacting bodies and *vice versa*. Going to the extremes of the size scales, it becomes even more apparent. For example, the quantum number makeup, spectroscopy, and even a further imaging (like [1]) of a simple Quantum Mechanical system such as a Hydrogen atom corresponds to a single kind of Electromagnetic interaction. The following scattering experiments allow us to learn the structures of different materials, which are invisible to the eye.

Similarly, for strong interactions, the structure of protons is used to improve our understanding of basic phenomena emerging from the existence of strong interactions. Beyond that, it provides a stepping stone to exploring properties of other interactions such as Electroweak and physics Beyond the Standard Model that are only accessible in scattering of elementary particles such as quarks and gluons.

1.2 Strong interactions

The strong interaction has important consequences for our daily lives as it holds quarks and gluons together to form hadronic matter. A common example of hadrons are protons and neutrons, also known as "nucleons"; they make up the nucleus of each atom inside and around us. The strong interaction gives rise to the nuclear force that binds those nucleons when holding each nucleus of all the matter seen in and around us together.

Unlike gravitational and electromagnetic interactions, strong interactions can not be described in classical macroscopic physics in terms of forces as in Newtonian mechanics. This has to do with the fact that the charge "color" that governs the presence of the strong interaction is "screened" to not be visible on distances beyond the nuclear sizes. It is only on scales of 10^{-15} m and smaller that the effects are present and are subject to the laws of the Quantum world.

1.2.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is a relativistic field theory that describes the strong interaction at a fundamental level. The description is performed in terms of the gluon gauge field and fermion fields of quarks of different flavors (up, down, strange, charmed, bottom and top). The dynamics of the interaction between those fields can be encoded in terms of the Lagrangian of QCD theory, which looks like:

$$\mathcal{L}_{\text{QCD}} = \sum_{f=1\ldots N_f} \overline{\psi}_f \left(i \sum_{\mu=0\ldots3} \gamma^\mu \mathcal{D}_\mu - m \right) \psi_f - \frac{1}{4} \sum_{\substack{\mu,\nu=0\ldots3 \\ a=1\ldots8}} (G^a_{\mu\nu})^2 \tag{1.1}$$

where $D_\mu = \partial_\mu - ig \sum_{a=1\ldots8} G^a_\mu \Lambda_a/2$ and $G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g \sum_{b,c=1\ldots8} f^{abc} A^b_\mu A^c_\nu$. Formally, knowing the Lagrangian of the theory should allow us to make predictions for all

[2] Known as Rutherford's gold foil experiment.

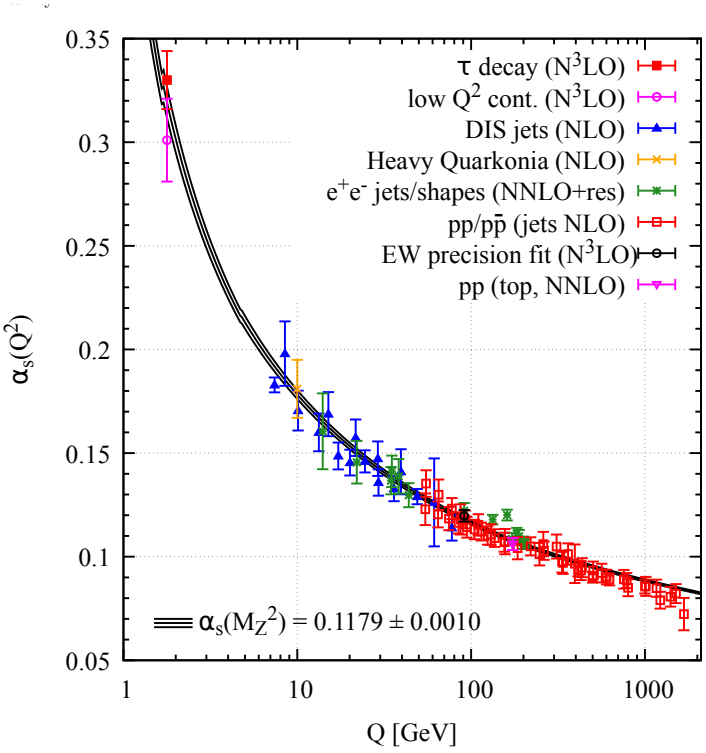


Figure 1.1: A global fit of the running $\alpha_s(Q)$ to the world data (plot from [2]).

the observable phenomena. Yet, with the methods we have today, there are some extreme complications when attempting to derive any of those results from first principles despite knowing the general structure of the theory. For example, in regards to scattering problems, the most powerful method of "perturbation theory" relies on our ability to approximate the full interaction in the field theory as an infinite sum of amplitudes of different processes. The processes are ordered in discrete powers of the coupling constant g that also signifies the number of discrete "small" interactions between otherwise free fields (particles). If the coupling strength g is sufficiently small, that sum should converge to the amplitude of the full physical process. What happens in practice is that the g is in many cases not small, which makes perturbation theory inapplicable.

As it turns out, gluons interact with themselves (via the "$g\sum_{b,c=1\ldots8} f^{abc}A^b_\mu A^c_\nu$" term) that makes the coupling g in the perturbative theory change with the scale of the process in such a way that the coupling increases at large distances (low energies), and it decreases at small distances (high energies). The dependence of the strong coupling constant $\alpha_s \equiv \frac{g^2}{4\pi}$ on the energy scale Q is shown in Fig. 1.1. The gluon self-interaction in QCD also leads to formation of gluon flux tubes between color charges which gives rise to the phenomenon of color confinement, which makes strongly-interacting elementary particles such as quarks, antiquarks and gluons unable to exist as free particles. Instead, only color-singlet (this is an analog of neutral electric charge) composite particles called hadrons can exist as a low energy system. The phenomena of the small coupling at high energy – Asymptotic freedom – as was already mentioned, allows for application of perturbation theory, which theoretically justifies our ability to probe the behavior of partons (quarks, antiquarks and gluons) inside hadrons.

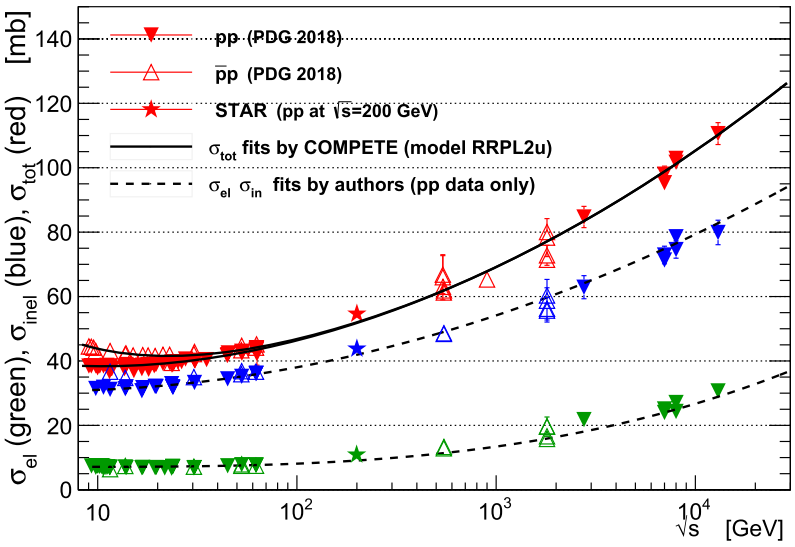


Figure 1.2: Dependence of the elastic, inelastic and total cross sections on the collision energy $\sqrt{s}$ for proton-proton and proton-antiproton interactions (plot from [3]).

1.2.2 Strong interactions in hadron-hadron collisions

Hadron-hadron collisions can occur via the following processes [4]:

- elastic scattering process
- inelastic scattering processes
  - single diffractive process (SD)
  - double diffractive process (DD)
  - non-diffractive process (ND)

There is an interference between the double diffractive and the soft non-diffractive processes, so from a Quantum Mechanical standpoint they are not separable. However, they are separated in the implementation of the Monte Carlo "Event Generators" such as PYTHIA [5]. In fact, it is a common thread of our state-of-the-art understanding of nature where a complex continuous process is well described by discrete steps coded in the event generators.

This work focuses entirely on the hard non-diffractive process, in which constituent partons experience hard scattering, which is a scattering with momentum transfer $Q \gg \Lambda_{QCD}$ (or equivalently $\alpha_s(Q^2) \ll 1$). The cross section for those processes at an energy of $\sqrt{s} = 200$ GeV is on the order of 30 mb, which is a sizable fraction of the total cross section that is just above 50 mb (as shown in Fig. 1.2).

A schematic view of a hard scattering event is given in Fig. 1.3. A general breakdown of the Event Generator's point of view would consist of the following components:

- Initial state of hadrons
- Beam Remnants

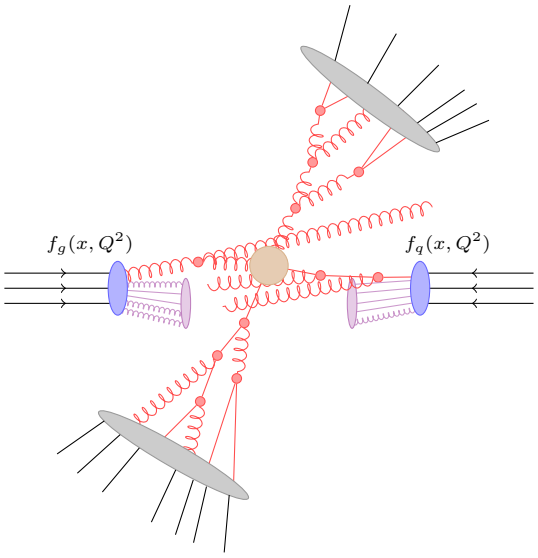


Figure 1.3: Schematic representation of an inelastic hard non-diffractive proton-proton collision: two partons from the incoming protons carrying a longitudinal momentum fraction x experience a hard scattering at the momentum scale Q, radiate in the initial and final state, and all outgoing partons hadronize. Multiple Parton Interactions that often occur are not shown for clarity.

- Initial State Radiation (ISR) interleaved with Multiple Parton Interaction (MPI)
- The hard collision of partons
- Final State Radiation (FSR)
- Hadronization

An example of an actual simulated hard scattering event is shown in Fig. 1.4 and allows one to get a sense of the complexity of those events.

Factorization and the proton structure

The asymptotic freedom implies that high transverse momentum products of hadron-hadron collisions would come from collision of individual partons and not hadrons as a whole, and that interaction can be calculated perturbatively. This motivates a formalism known as Factorization, which is a set of theorems that are used for separating the short distance perturbative physics of the initial and final states from the long distance non-perturbative physics. Factorization has been mathematically proven for processes such as Deep Inelastic Scattering (lepton + hadron $\rightarrow$ lepton$'$ + X), and Drell-Yan (hadron + hadron $\rightarrow \gamma/Z + X$) [6].

The physical principle behind Factorization can be understood via the following reasoning: Consider a scattering between a parton and a proton. In the center-of-mass reference frame the proton is contracted in the direction of the collision axis. As the energy increases, the lifetime of a partonic state inside the proton is dilated – the travel path for the incident parton is getting smaller. In this limit the other partons inside the proton will appear to

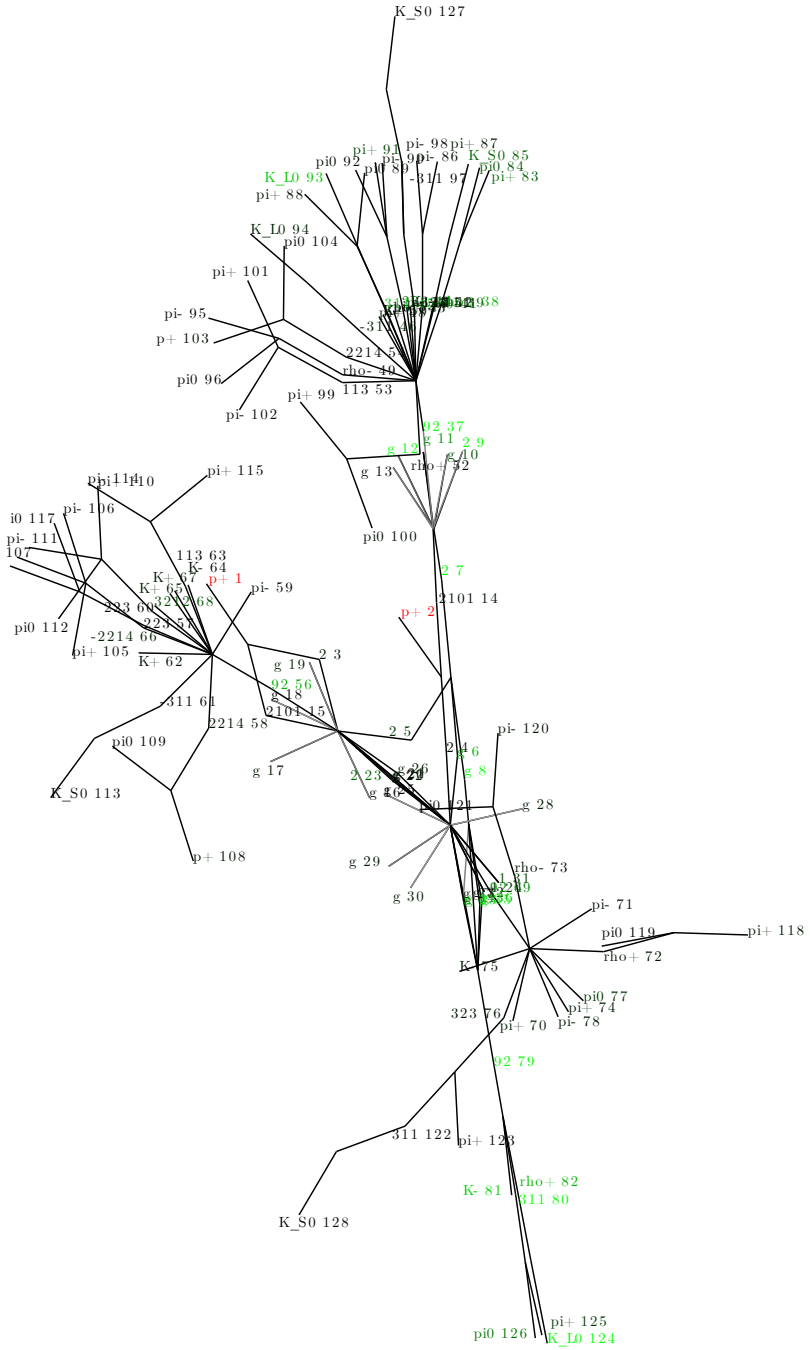


Figure 1.4: A graph (similar to a Feynman diagram) showing an example of a hard scattering event generated by the HERWIG 7 Event Generator for protons colliding at the energy $\sqrt{s} = 200$ GeV. The final state particles at the edges are pulled according to φ-direction of their momenta. The initial protons (with an undefined φ) are labeled in red. Green labels correspond to particles with high transverse momentum p_{T}.

be non-interacting. Each one of those can be then considered as carrying a fraction x of the full proton momentum. The "frozen" state means that the cross section for interaction of the parton with that state can be computed by a summation of probabilities instead of amplitudes. This leads to following expression:

$$\sigma(\alpha_{\mathrm{s}}(\mu_{\mathrm{R}}), \mu_{\mathrm{R}}, \mu_{\mathrm{F}}) = \\ = \sum_{a,b} \int \underbrace{f_a(x_1, Q^2) f_b(x_2, Q^2)}_{\text{proton structure}} \underbrace{d\hat{\sigma}_{a+b\to \text{jet}+X}(x_1, x_2, Q^2)}_{\text{hard process+PS+Had.}} \, dx_1 \, dx_2 + \mathcal{O}\left(\frac{1}{Q^2}\right) \quad (1.2)$$

where $\sigma(\alpha_{\mathrm{s}}(\mu_{\mathrm{R}}), \mu_{\mathrm{R}}, \mu_{\mathrm{F}})$ is the cross section for the hadron-hadron interaction, and $f_a(x_1, Q^2)$ is the Parton Distribution Function for a parton of kind a with momentum x_1 at the momentum scale Q. The $\hat{\sigma}$ is known as the cross section for the parton scattering process and is assumed to be calculable. The Parton Distribution Functions (PDFs) are non-perturbative and have to be measured, although there is some recent progress in calculating them numerically in the lattice QCD approach [7]. Another important property of PDFs is that they are universal, meaning that once they are measured for a hadron in a one specific process at a specific energy Q, they can be used to predict a different process involving the same hadron with a well defined way to perform evolution to another energy.

Hard scattering of partons

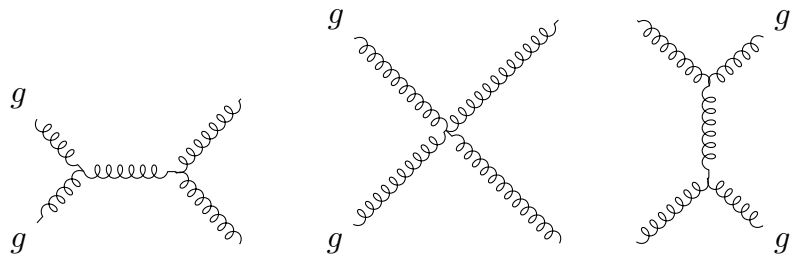
If one is to look at Fig. 1.3 as if it was a Feynman diagram, the hard scattering would be the blob with the highest momentum exchange involved. At the leading order, that blob would correspond to a 2-to-2 (two incoming into two outgoing partons) process represented by diagrams from Fig. 1.5.

The cross sections $\hat{\sigma}_{\text{2-to-}N}$ for the parton-parton scattering itself is well described by perturbative QCD. Those results are relatively complex, but many calculations are available: standalone and included as a part of the Event Generators. A question that is simpler to analyze, but still of importance, is what is the connection between the kinematics of the initial state part**ons** and the final state part**icles**.

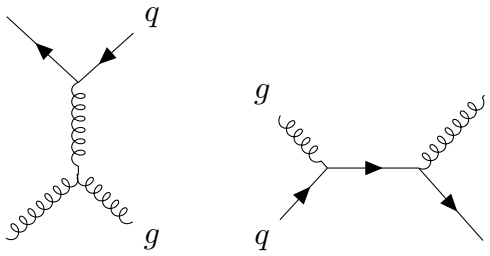
Consider two relativistic protons with momenta $\pm\sqrt{s}/2$ colliding in their center-of-mass frame and experiencing a hard parton scattering. Say the two scattering partons carry fractions of their respective protons longitudinal momenta x_1 and x_2. Writing the energy-momentum conservation law for this 2-to-N process:

$$\begin{pmatrix} x_1 \frac{\sqrt{s}}{2} \\ 0 \\ 0 \\ x_1 \frac{\sqrt{s}}{2} \end{pmatrix} + \begin{pmatrix} x_2 \frac{\sqrt{s}}{2} \\ 0 \\ 0 \\ -x_2 \frac{\sqrt{s}}{2} \end{pmatrix} = \sum_i \begin{pmatrix} E_i \\ p_{x;i} \\ p_{y;i} \\ p_{z;i} \end{pmatrix} \quad (1.3)$$

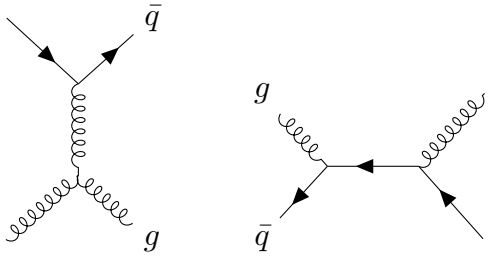
where index i enumerates the particles produced from the parton scattering. Solving for x_1 and x_2 one gets:



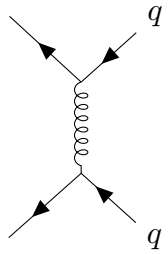
(a) gg scattering



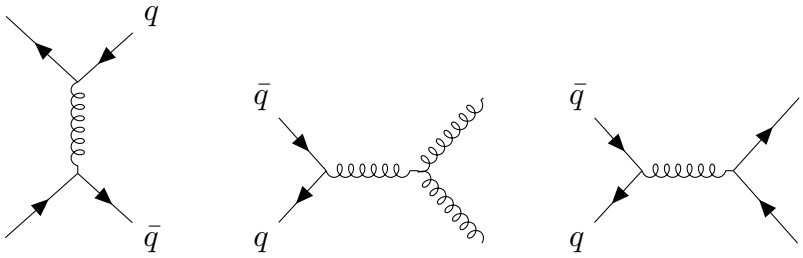
(b) qg scattering



(c) $\bar{q}g$ scattering



(d) qq scattering



(e) $q\bar{q}$ scattering

Figure 1.5: Feynman diagrams corresponding to the leading order parton scattering. Diagrams that can be achieved via cross symmetry (i.e., changing the topology by rearranging incoming and outgoing "legs") are omitted.

$$\begin{cases} x_1 = \dfrac{1}{\sqrt{s}} \sum_i (E_i + p_{z;i}) = \dfrac{1}{\sqrt{s}} \sum_i \sqrt{m_i^2 + p_{T;i}^2} \exp(y_i) \\ x_2 = \dfrac{1}{\sqrt{s}} \sum_i (E_i - p_{z;i}) = \dfrac{1}{\sqrt{s}} \sum_i \sqrt{m_i^2 + p_{T;i}^2} \exp(-y_i) \end{cases} \tag{1.4}$$

where the definition of the rapidity y is given by

$$y \equiv \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) \tag{1.5}$$

These general equations are not exactly applicable to the final state hadrons. In order for colored partons to become colorless hadrons in the process of hadronization, there must be an exchange of color between the partons outgoing from the scattering and the beam remnants. The mediators of that exchange will, along with color quantum numbers, carry some momentum between the systems. Thus, the assumption of energy-momentum conservation for the system of the two colliding partons is slightly invalidated for a process that includes hadronization.

In application to $2 \to 2$ scattering, in the limit of $m_i \ll p_{T;i}$, one gets simpler expressions[3]:

$$\begin{cases} x_1 \simeq \dfrac{p_{\mathrm{T}}}{\sqrt{s}} \left(\exp(\eta_1) + \exp(\eta_2) \right) \\ x_2 \simeq \dfrac{p_{\mathrm{T}}}{\sqrt{s}} \left(\exp(-\eta_1) + \exp(-\eta_2) \right) \end{cases} \tag{1.6}$$

where the definition of the pseudorapidity η is given by

$$\eta \equiv \frac{1}{2} \log \left(\frac{p + p_z}{p - p_z} \right) \tag{1.7}$$

This shows that by being able to measure the 3-momenta of the outgoing partons one can recover the initial momentum fractions of the partons inside the protons. In combination with a jet production rate measurement, this and the factorization theorem from Eq. (1.2) along with a theoretical calculation of $\hat{\sigma}_{a+b \to \mathrm{jet}+X}$ would allow one to probe the Parton Distribution Functions.

1.2.3 Radiation and Jets

It is a well known effect that accelerating charges produce radiation. This is true for electrically charged objects that radiate photons, but also true for particles such as partons that carry the strong "color" charge, and they radiate gluons. As a part of the same effect, off-shell gluons may convert into quark-antiquark pairs. A commonly known property of the radiation from hard acceleration is that the radiated particles prefer to have low transverse momenta with respect to either the direction of the initial charge movement or the direction of the final charge movement. These two processes correspond to the ISR and FSR (initial state and final state radiation, respectively). In each class, the radiation can be subdivided into

[3]This can be also applied to jets that use "E-scheme" for recombination.

soft radiation, which doesn't have much of a preferential direction, but also doesn't carry much energy, and hard radiation which is collimated in one of the two cones pointing in the direction of the charged particle's initial or final momenta.

Jets, as physical objects, are a collection of radiation formed by the FSR in the hard collision. As was previously mentioned, the hadronization process results in momentum exchanges between the color-connected partons. This causes changes to the observed distributions of outgoing hadrons, but the resulting jet-like structures still persist in sufficiently hard events, in which the hadron jets are experimentally observed.

The significance of jets is that by identifying these radiation patterns in the events one is able to deduce information about the outgoing partons from the scattering. The physics behind the radiation and especially hadronization is complicated, but the laws of momentum and charge conservation allow one to make some reasonable assumptions. For example, jets tend to have a momentum that is strongly correlated to the momentum of its original parton.

To make use of jets, one needs a practical way to identify them in the sets of detected particles. Such a procedure, commonly referred to as the "jet algorithm", is designed to define jets that are best at reproducing select characteristics of the outgoing partons. To that end, they have to take into account the general properties of the processes at hand (listed in Section 1.2.2) and requirements imposed by the theoretical calculations.

One such algorithm, the "anti-k_{T}" jet algorithm [8], is the most popular at the time of this writing and has some provisions to reduce the impact of low-p_{T} particle backgrounds, Underlying Event and pile-up on the jet thrust axis direction. It is also "infrared-safe" and "collinear-safe" with respect to perturbative calculations in all orders of α_s.

1.3 Measurements of the jet production cross sections

Jets are pervasively used in modern High Energy Physics, and it is a vast field to cover. This section will focus on topics related to jet cross sections.

1.3.1 Observation of jets in e^+e^- annihilation

At the dawn of jet physics, jet structures were commonly studied in terms of event shape quantities such as sphericity S and thrust T determined with respect to a selected jet axis. Event shape quantities measured near resonances such as Υ were best described by a model based on an intermediate state that decayed into three gluons. Those would have a drastically different event shape compared to the $e^+e^- \to q\bar{q}$ channel that is dominant at off-resonance energies. That observation would serve as first direct evidence for existence of gluons.

Jets produced in e^+e^- annihilation are at a constant energy, which allowed theorists to tune the FSR and many of the aspects of hadronization (except for the effects related to color connections to the beam remnants).

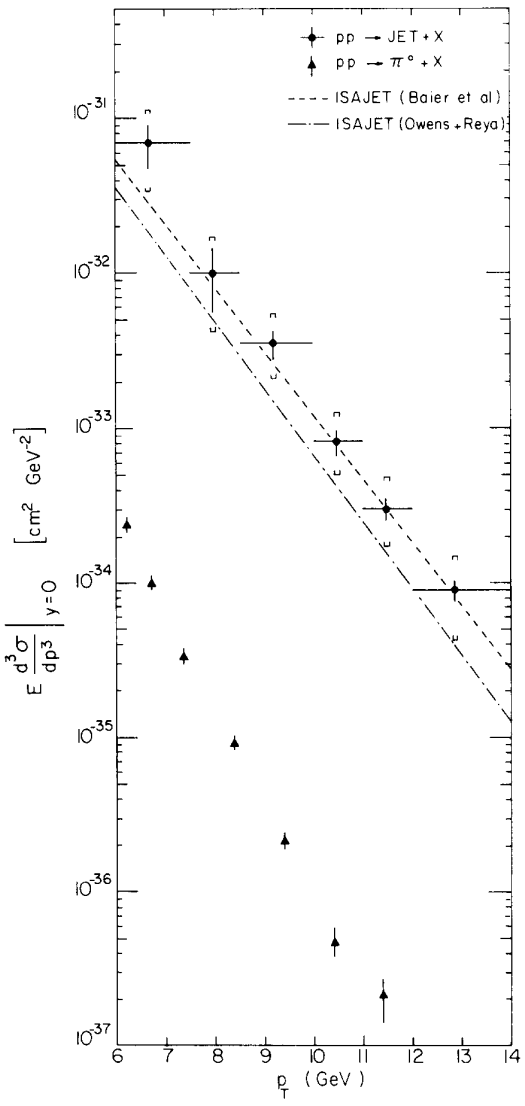
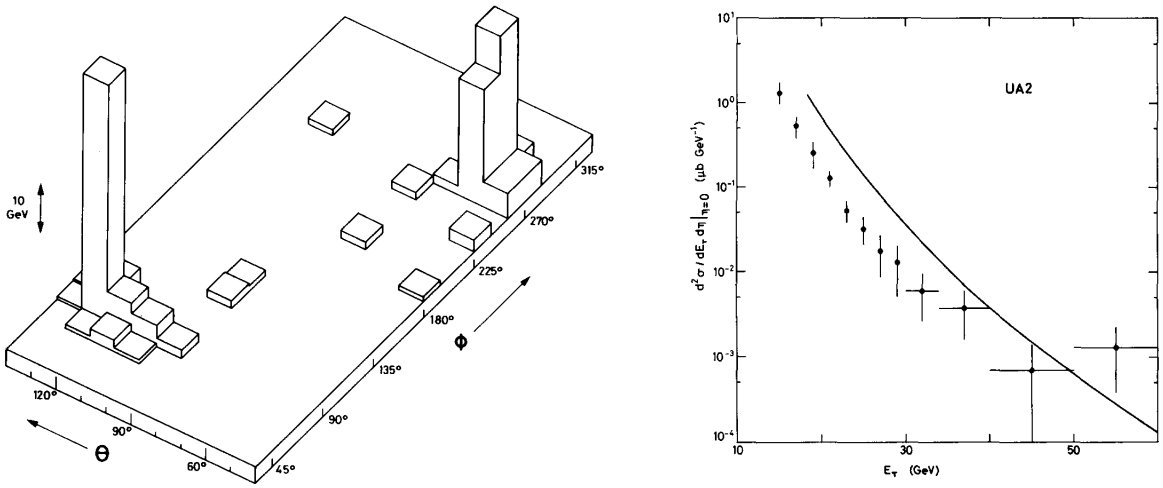
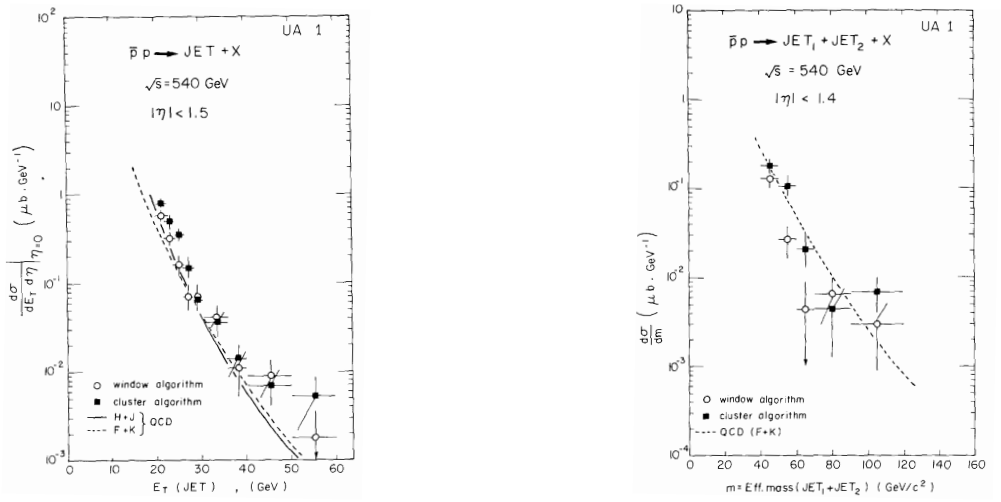


Figure 1.6: Inclusive jet cross section for jet production in proton-proton collisions measured by the AFS experiment at the ISR collider at $\sqrt{s} = 63$ GeV [9]. A cross section for π^0 production is provided for comparison. A Monte Carlo simulation was used to map observed transverse energy to the transverse momentum of the original constituents.



(a) (b)

Figure 1.7: First measurement for jet production in proton-antiproton collisions by the UA2 experiment at the CERN S$\bar{p}$pS collider at $\sqrt{s} = 540$ GeV [10]. Jets were defined as clusters of adjacent calorimeter cells that contain more than a fixed fraction of the energy deposited in the calorimeter. (a) An example event with two back-to-back jets with $M = 140$ GeV. (b) An inclusive jet cross section (17% luminosity uncertainty not shown) that was corrected for detector effects using a bin-by-bin factors estimated from simulation. The solid line showed a QCD prediction that was available at the time.



(a) Inclusive jet cross section (b) Di-jet cross section

Figure 1.8: Jet cross section measurements by UA1 experiment at S$\bar{p}$pS for proton-antiproton collisions at $\sqrt{s} = 540$ GeV [11]. Results were derived from the data with two different definitions for jets. Detector effects were implemented as corrections to the E_T scale, as well as a global multiplicative factor on $\mathrm{d}\sigma/dE_{\mathrm{T}}$.

1.3.2 Jet production cross sections in hadron-hadron collisions

Historical observations

While some fixed target experiments such as NA5 (at $\sqrt{s} = 24$ GeV) claimed some observation of jets ([12, 13]), historically, it was not until the first hadron colliders were built that jets were reliably observed. This has to do with the fact that jet production requires a sufficiently high energy to achieve collimation of the radiation cone. The hadron colliders were able to achieve greater collision energies thanks to the reduced Synchrotron radiation for protons that are much heavier than electrons.

The first evidence for collimated jets was reported by the Axial Field Spectrometer (AFS, also known as R807) detector at the Intersecting Storage Rings (ISR) hadron collider in pp collisions at $\sqrt{s} = 63$ GeV by studying event shapes in its hadronic calorimeter [9] and in its tracker [14]. The first unambiguous evidence for back-to-back 2-jet structures resulting from hard parton scattering were observed at the UA2 experiment at CERN's S$\bar{\text{p}}$pS collider [10] (Fig. 1.7a). It is notable that since those early experiments, the measurements of inclusive jet cross sections (see Figs. 1.6 and 1.7b) were customary, although the practice of having a rigorous definition of a jet using jet algorithms was not formalized until some time later.

A later publication by the UA1 experiment [11] at the same S$\bar{\text{p}}$pS collider emphasized a need to unambiguously define a "jet" before trying to do quantitative measurements of them. This was the first measurement featuring a jet algorithm based on grouping calorimeter cells according to the metric

$$\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{1.8}$$

Experiments at the Tevatron proton-antiproton collider, at much higher energies of $\sqrt{s} = 1.8$ and 1.96 TeV, made progress by providing jet cross section measurements suitable for quantitative interpretations. Measurements at the D0 experiment [15–19] and CDF experiment [18, 20] relied mostly on cone algorithms for the jet reconstruction. Major improvement was made in the procedures implementing corrections for the detector effects. Advances in Monte Carlo simulations at the time allowed implementation of the first theoretical corrections for non-perturbative effects such as hadronization, ISR and MPI to facilitate comparisons with the NLO pQCD calculations. With a maximum jet p_{T} probed at ≈ 600 GeV, the kinematic x-reach of measurements at Tevatron was $x \lessapprox 0.6$ (for $\eta_1 = \eta_2 = 0$ in Eq. (1.6)).

The first measurement of the inclusive jet cross section at STAR [21] (at $\sqrt{s} = 200$ GeV) was using the CDF mid-point cone algorithm with $R = 0.6$ and split fraction 0.5. Unlike modern measurements, that one did not apply a correction for the detector effects contributing a difference between the observed and true jet transverse momentum. Present thesis improves in that regard by employing an unfolding procedure using a simulated detector response (Chapter 5).

In the era of the Large Hadron Collider, the new generation of detectors along with enhanced analysis procedures allowed for precise measurements of jet cross sections.

1.3.3 Modern uses for inclusive jet spectra

One benefit of using hadron-hadron collisions, compared to the e^+e^- annihilation and Deep Inelastic Scattering, is that they probe the gluon PDF in the leading order of pQCD through

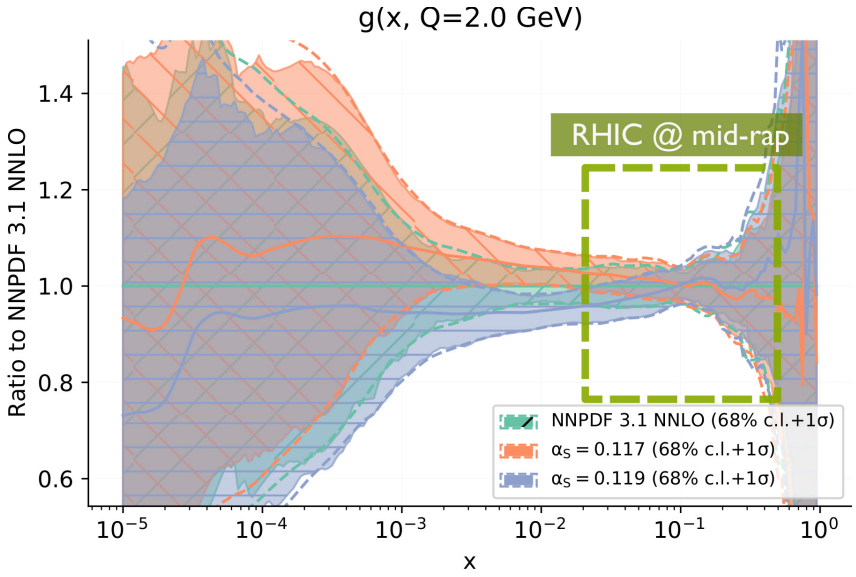


Figure 1.9: Parton Distribution Function of gluons inside protons as a function of x according to the fit from the NNPDF collaboration. Derived from [25].

gg, qg and $\bar{q}g$ scattering (diagrams shown in Figs. 1.5a to 1.5c).

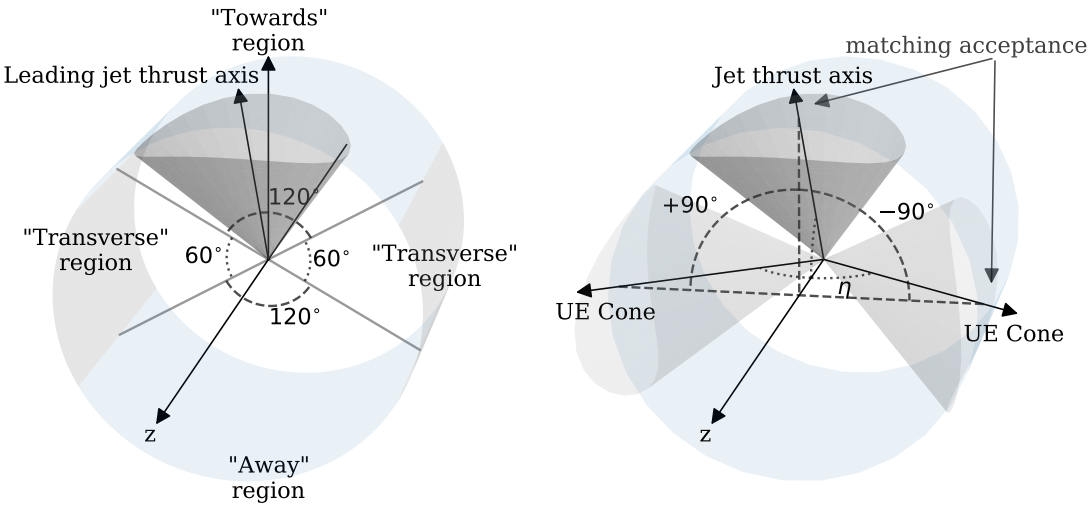
A recent paper [22] describes the most recent (at the time of this writing) measurement of the differential inclusive jet cross section done by CMS for pp collisions at $\sqrt{s} = 13$ TeV. It presents an analysis of its cross section measurement using Standard Model predictions extended by a set of three dimension-6 operators corresponding to a color-singlet exchange between two quark lines. Model parameters for the PDF set and the three coefficients for the Beyond Standard Model physics operators were extracted using a method for reweighting PDF sets using their Hessian error vectors [23] and using the full fits. While the measurement did provide an improvement in PDF fits over the standard CT14nlo set [24], the study should be able to benefit from a dataset that is directly sensitive to gluons at high-x, but does not include any effects of BSM physics effects (e.g., by instead conducting a measurement at a lower value of $\sqrt{s}$).

1.3.4 Inclusive jet cross section at STAR

A measurement of the inclusive jet cross section would allow us to reach $x \lessapprox 0.55$ and $x \lessapprox 0.32$ for $\sqrt{s} = 200$ and 510 GeV, respectively. The measurements at different energies can well complement each other, especially if measured in the same year, with correlated systematic uncertainties. This thesis focuses on the measurement at $\sqrt{s} = 200$ with the goal of providing a better constraint on the high-x gluon PDF, which, as seen in Fig. 1.9, is relatively poorly constrained by current data.

Underlying Event subtraction

The events including physical processes listed in Section 1.2.2 contain particles without any particular labels with respect to which process produced them. In fact, because of



(a) "Region" method (b) "Cone" method

Figure 1.10: Two definitions of the transverse regions to sample the Underlying Event activity with respect to jet's thrust axis. The other possibly present jets from the event are omitted for clarity.

hadronization effects, even Event Generators can be quite ambiguous about the event history. Some particles in the measured event may be coming from jets, and some from the MPI, or the ISR, or the beam remnants.

As with jets, we can rely on event topology to perform some separation for the contributions. The traditional method to do this, originally used to tune MPI models of the Monte Carlo generators using CDF data on the Underlying Event, involves defining two regions in azimuthal angle that are transverse with respect to the leading jet as shown in Fig. 1.10a. The density of particles produced in that phase space can be nominally assigned to the Underlying Event. As models for the MPI allow for a fluctuating number of additional soft scatterings, the energy density of the Underlying Event is expected to fluctuate between the events.

Previous differential inclusive jet cross section measurements at all of the experiments would often be compared to a cross section predicted by NLO pQCD and corrected using multiplicative factors for the UE and hadronization derived from a Pythia simulation. The present measurement will use a data-driven method from [26] by applying a correction for the UE from the measurement itself on an event-by-event basis. It has been successfully adopted by the polarized measurements at STAR. It suggests estimating the contribution of the UE to the jet p_{T} as

$$\Delta p_{\mathrm{T}}^{\mathrm{UE}} = \rho_{\mathrm{UE}} \cdot A_{\mathrm{jet}} \tag{1.9}$$

where $\rho_{\mathrm{UE}} \approx 1.2$ GeV/radian/$[\Delta\eta]$ is the momentum density of the UE per unit $\mathrm{d}y\,\mathrm{d}\varphi$ of phase space, and A_{jet} is the jet area (see [27] for the definition), which for the isolated anti-k_{T} jets is simply $\pi R^2 \approx 1.13$. For "overlapping" jets the overlap is split according to the magnitudes of the jets' momenta. The suggested way to estimate the ρ_{UE} is to sample it for each jet individually in the transverse jet cones with $R_{\mathrm{cone}} = R_{\text{anti-}k_T}$ pointing at $\varphi_{\mathrm{cone}} = \varphi_{\mathrm{jet}} \pm 90°$ and $\eta_{\mathrm{cone}} = \eta_{\mathrm{jet}}$ (as pictured in Fig. 1.10b). The "correction" is then

applied by simply subtracting the jet's $\Delta p_{\mathrm{T}}^{\mathrm{UE}}$ from its p_{T}. Note that the cross section differential in this new jet p_{T} has a different definition now, which has to be respected in future interpretations (e.g., when comparing against the predictions of the Monte Carlo generators).

1.4 Thesis outline

The present thesis describes a measurement of the inclusive jet cross section production at $\sqrt{s} = 200$ GeV. It is performed at the STAR experiment, which will be discussed in Chapter 2. The description of the specific dataset, the specific steps taken to extract the jets from it and some simple jet observable will be discussed in Chapter 3. Chapter 4 will discuss a computer simulation needed to estimate the response of the detector to the jets. Approaches used to correct the data for the detector response will be discussed in Chapter 5 with some generality. The end procedure for the measurement and its result will be discussed in Chapter 6. The applicability of the result is then evaluated using a QCD theory framework in Chapter 7. Conclusions will be made in Chapter 8.

Chapter 2

Experimental setup

The description of the present measurement will rely on data recorded for proton-proton collisions produced at an intersection of proton beams at the RHIC Accelerator Complex described in Section 2.1 and partially in Appendix A. The events with jets are registered by a general-purpose detector – "Solenoidal Tracker at RHIC" (STAR). The components (i.e., subdetectors) of STAR that are relevant to the present measurement are described in Section 2.2.

2.1 Relativistic Heavy Ion Collider

The Relativistic Heavy Ion Collider [28] (RHIC) has two circular beamlines capable of accelerating vertically polarized protons. There are six interaction points where a two counter-rotating beams cross. The beams are commonly referred to as "Blue" and the "Yellow", respectively. The beam rings are 3834 meters[1] in circumference and employ superconducting magnets. The nominal magnetic field for 250 GeV protons in the arc dipole magnets is 3.4 T at a current of 5 kA and temperature below 4.6 K [29].

The polarized proton beams are first produced at energy of 35 keV and polarization $70 - 75\%$ in an Optically Pumped Polarized H^- Source (OPPIS) [30]. Then they are accelerated to 750 keV in a Radio Frequency Quadrupole low energy accelerator and then further to 200 MeV in the linear accelerator and further, again, in a Booster to 1.5 GeV before entering the Alternating Gradient Synchrotron serving as a booster accelerating the beam up to the RHIC injection energy of about 24 GeV. Finally, the beam rings of RHIC are where the beam is injected in bunches and accelerated to the final proton beam energy of typically 100 or 255 GeV. During the beam acceleration two helical magnet setups called "Siberian snakes" are rotating proton spins to avoid depolarizing resonances, which allows achieving polarization of $45 - 55\%$ at flattop. The nominal configuration for running physics with proton-proton collisions entails a 120-bunch beam in each RHIC ring. Each bunch carries on the order of 10^{11} protons and is able to have an independent polarization direction. Some 10 consecutive bunches remain unfilled to provide the "abort gap" required by the hardware that facilitates planned beam dumps.

[1] At the speed of light, one rotation takes about 12.8 μs

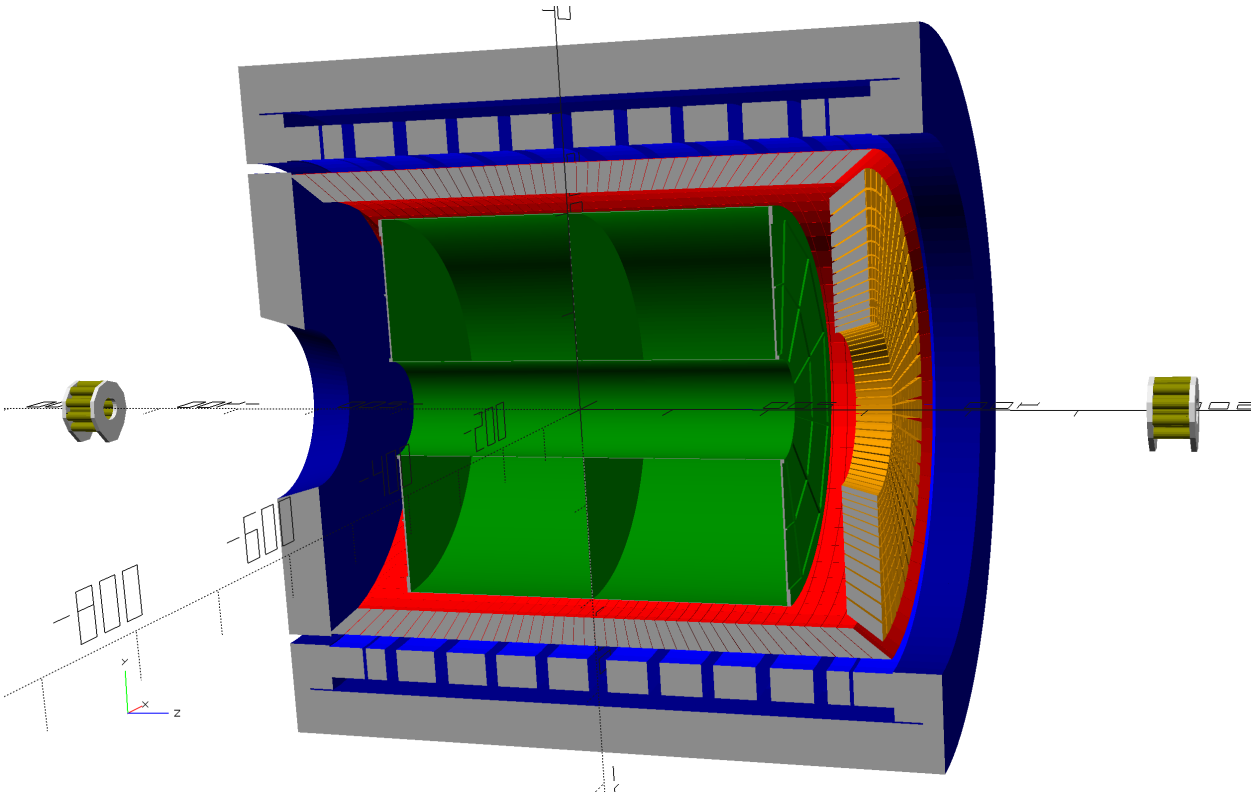


Figure 2.1: Schematic image for the STAR detector. A vertical cross section of the major components is presented: TPC (in green), BEMC (in red), EEMC (in orange) and the Magnet (blue). VPD (in olive) are situated forward and backward along the beam.

In year 2012, at the STAR interaction region, individual bunches were on average polarized along the direction of their momenta (after being rotated from and to vertical beam polarization using the "Spin Rotator" magnets). The effects of this polarization for unpolarized measurements are cancelled by averaging over all bunches, since they were polarized in an alternating up and down spin pattern. The effects remaining after the averaging are negligible due to the small magnitude of the spin asymmetries, which in turn results in relatively small spin-dependent effects on observables such as jet production cross sections. For this reason, the polarization of the beams is not considered throughout this thesis.

2.2 The STAR detector

2.2.1 Solenoidal Magnet

The STAR detector uses a large room temperature solenoidal electromagnet (shown in Fig. 2.1 in blue) that creates a nearly uniform magnetic field inside of its barrel. The vector $\vec{B}$ of the magnetic field is pointing along the z-axis, collinear with the direction of the proton beam momenta, which in turn makes charged particles bend in the azimuthal direction.

Measurement of the curvature of the tracks in the charged particle tracker allows one to estimate the charged particle's transverse momentum.

During Run 12 operation, the magnetic field inside the coils was pointing in the -z direction and had a nominal flux density of ~ 0.5 T, which corresponds to a current of about 4.7 kiloamperes running through the main coil [31].

2.2.2 Time Projection Chamber (TPC)

The TPC can be roughly described as a barrel 4.2 meters long and 4 meters in diameter (shown in Fig. 2.1 in green) filled with a gas[2] [32]. Whenever a charged particle passes through the gas volume of the TPC it will collide with the gas particles and create ionization. A uniform electric field created inside the TPC makes electrons from the ionization to drift away from the center membrane towards the endcap sides of the barrel. In the end, they approach the anode wires where they create an avalanche ionization. An amplitude of the image charge from that ionization is then read out from the sensitive pads. Measurement of the drift time is used to determine the distance from the initial ionization to the TPC endcaps, thus providing the measurement for the z coordinate of a point on the track. The pads are arranged in 45 rows so that the trajectory of a given relatively straight, outgoing track can be potentially constrained at 45 positions (at different "reconstruction layers").

2.2.3 Barrel Electromagnetic Calorimeter (BEMC)

The second most important subdetector (after the TPC) for mid-rapidity measurements at STAR is called BEMC [33] (Fig. 2.1 in red). Located just outside the TPC, after the Barrel Time Of Flight detector, the Barrel Electromagnetic calorimeter has a near 2π azimuthal coverage in the pseudorapidity region $|\eta| < 1$ for collisions at the center of the barrel (vertex position on the beam at $v_z = 0$). East and West halves are assembled out of 60 calorimeter modules, making it 120 in total for the whole detector. Each module is segmented into 2 rows of 20 towers projective towards $z = 0$. The size of each tower is $\Delta\eta \times \Delta\varphi = 0.05 \times 0.05$. The internal construction of the BEMC calorimeter module is one of a sampling calorimeter built with 20 layers of lead and 21 layers of plastic scintillator that are sandwiched together in an assembly that is about 26 cm deep, corresponding to 20 radiation lengths at $\eta = 0$ and to containment of the 60 GeV EM showers.

The Electromagnetic Calorimeters (EMCs) are crucial to jet measurements at STAR. Many of the hadrons such as neutral pions π^0 that are formed in hadron jets have a natural decay to photons, which, like any neutral particle, would be invisible to the TPC. The EMCs allow one to measure energy of those high energy photons and thus reconstruct a significant fraction of the jet's energy (about 30%). Another important property of the EMCs is that they are fast detectors and, as it will be discussed in Section 2.2.7, can be used to implement trigger conditions to select events with high fraction of jets.

[2]The gas mixture used is P10 – 10% methane and 90% argon

2.2.4 Endcap Electromagnetic Calorimeter (EEMC)

A second electromagnetic calorimeter called EEMC [34] (shown in Fig. 2.1 in orange) extends acceptance coverage with the range $1.086 < \eta < 2.00$. Compared to the BEMC, the EEMC towers are also projective towards the origin of STAR, but have a bigger size spanning $\Delta\varphi = 0.1$ in azimuthal angle and $\Delta\eta$ from 0.057 to 0.099. Each tower has 23 layers of absorber made of calcium-loaded lead laminated with stainless steel and 24 layers of plastic scintillator along the z axis, the total thickness of the assembly represents about 21 radiation lengths. The energy resolution is $\delta E/E \simeq 16\%/\sqrt{E}$. Using electronics identical to the electronics used for the BEMC tower readout and triggering allowed for good integration with BEMC for the jet triggers.

A special feature of the EEMC design is that it includes a fine granularity Shower-Maximum Detector (SMD) aimed at discriminating between single photons and photon pairs from π^0 and η^0 decays. Ability to resolve individual photons and measure the angle between them comes essential for the calibration of the EEMC which uses it to relate the observed energy to the well-known pion mass peak.

2.2.5 Zero Degree Calorimeters (ZDC)

Zero Degree Calorimeter [35, 36] is a detector subsystem consisting of two hadron calorimeters situated behind the East and West "DX" dipole magnets, or about 20 meters from the center of STAR along the beam line. Such positioning allows the ZDC to detect outgoing neutrons from the collision at a small angle with respect to the colliding beams, as all charged particles are deflected by the magnetic field of the DX magnets.

2.2.6 Vertex Position Detectors (VPD)

Vertex Position Detector [37] is implemented as two identical assemblies surrounding the beamline, located at both the East and West side at a distance of 5.7 m from the center of STAR (shown in Fig. 2.1 in olive). This detector's acceptance covers about a half of the solid angle within a $4.24 \leq \eta \leq 5.1$ pseudorapidity range. The VPD has two roles in STAR. The first role is that it can serve as a trigger detector for collecting so-called "Min-Bias" data (the particles produced in a forward region will tend to have a small transverse momentum p_{T} and, as such, indicate a possibility of either a soft or hard event). The respective trigger condition requires at least one hit to be present on both sides and, additionally, a condition is placed on the difference between the East and West times of flight to select mostly events near the center of STAR ($|v_z| < 30$ cm), where STAR's overall acceptance was highest. The second role is to provide a measurement of the start time for the Time of Flight detector. For that, the signal arrival times for both East and West halves are used:

$$t_{\mathrm{start}} = (t_{\mathrm{East}} + t_{\mathrm{West}})/2 - \frac{L}{c} \tag{2.1}$$

This results in a precision of $\sim$ 80ps for $p + p$ collisions.

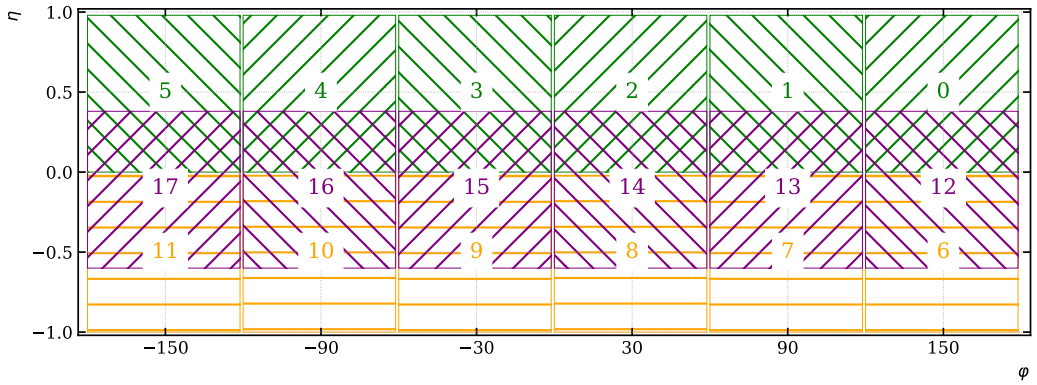


Figure 2.2: Map of the BEMC jet patch trigger patches in the η-φ plane. The jet patches covering only EEMC and those spanning across BEMC and EEMC are not shown here.

2.2.7 Trigger system

It would be impractical for the STAR detector to record events at each bunch crossing ($f \simeq 10$ MHz). Additionally, many important subdetectors in STAR are limited to operating at much lower rates (e.g. the TPC's operation was limited to ~ 1 kHz). A trigger system [38] implemented in the STAR detector addresses this problem. During the data taking, at each bunch crossing it uses information from fast detectors to make a decision whether an event of interest has occurred and whether it should be recorded for future analysis. At the time of Run 12 data-taking, the decisions for various triggers used information from the following detectors: ZDC, BBC, VPD, BEMC, EEMC and FMS. Information from the BBC and FMS detectors was not used in this analysis, and so they will not be discussed here.

The L0 triggering system uses Data Storage and Manipulation (DSM) boards [39] to define trigger conditions. Each DSM board has inputs and outputs implementing a programmable lookup table. The DSM boards are arranged in a tree having three levels with the lowest Level-0 DSM boards interfacing with the trigger detector digitizers and passing aggregated information to the Level-1 DSM boards. The information is further aggregated to a single Level-2 DSM board to make the decision on what trigger conditions have been satisfied. A device called Trigger Control Unit (TCU) takes output information from the DSM tree and decides if the trigger should be issued, and also implements trigger prescales. If the trigger is issued, the system issues a command to the required slow detectors to begin reading out. The information from the DSM boards is also recorded and can be used to perform offline checks on the data and on the integrity of the trigger system.

Jet Patch triggers

Triggers important for jet physics at STAR include a family of triggers called "jet patch" (JP) triggers [40]. Those rely on the information from both the BEMC and EEMC. The trigger decision is based on the summed energy values from the calorimeter towers in a square $\Delta\eta \times \Delta\varphi = 1 \times 1$ blocks of towers called "jet patches". Figure 2.2 shows the location of the

jet patches in the BEMC. Whenever the energy deposited in the jet patch would exceed a certain threshold (set in ADC counts), and the detectors participating in the data acquisition were ready to be read out (not reported as "busy" or "dead" state in the trigger system), and the prescaling mechanism accepted an event, the trigger system would issue a command to the DAQ for the event to be read out and recorded. The use of large jet patches allows capturing large samples of jets, however those samples will tend to exclude jets with low energy deposition in the electromagnetic calorimeters, and may further exclude some fraction of "wide" jets as those have a higher chance to deposit a significant fraction of their energy outside of their primary jet patch into an adjacent one.

Due to the steeply falling nature of the jet cross sections as a function of the jet transverse momentum p_{T}, a lower trigger threshold will generally result in a higher trigger fire rate. In order to capture the low p_T jets at a reasonable trigger rate and preserve as many as possible of the rare high p_T jets, three jet patch triggers with different thresholds and prescale factors were used: JP0, JP1 and JP2. The prescale factors were dynamically adjusted to adjust for varying collider luminosity to fit within the allocated bandwidth.

Chapter 3

Dataset processing and quality assurance

The STAR experiment is a complex system that undergoes year-to-year improvements, but is also subject to variations in its performance, including occasional technical failures. Any effect of the latter needs to be detected by evaluating data for anomalies.

This chapter describes the data set, procedures used to reconstruct jet events, and steps performed to select the parts of data that are of a good quality.

3.1 The Data sample

The data used in this analysis were collected during STAR operation for proton-proton collisions at $\sqrt{s} = 200$ GeV in the year 2012. The detector configuration in that year had the minimal amount of material in the way of mid-rapidity particles as there were no vertex detectors installed between the beam pipe and the TPC. There were also no significant operational issues, and the data were well reproduced by the simulation.

These data have been used in published measurements of

1. Azimuthal Single-Spin Asymmetries of Charged Pions in Jets [41, 42]
2. J/Ψ production cross section and its dependence on charged-particle multiplicity [43]
3. Underlying event activity [44]
4. Groomed Jet Substructure Observables [45]
5. Inclusive J/Ψ polarization [46]
6. Invariant jet mass [47]

The full dataset for the `pp200_production_2012` trigger configuration spans 1189 data-taking "runs" between 13038134 and 13075024 taken between February 7th and March 12th.

After an initial quality assurance selection done in [42], only 601 runs were deemed usable for mid-rapidity jet analyses. Those runs correspond to a total of 925 million of recorded events for all the various triggers, which corresponds to an integrated luminosity of 17 pb^{-1}.

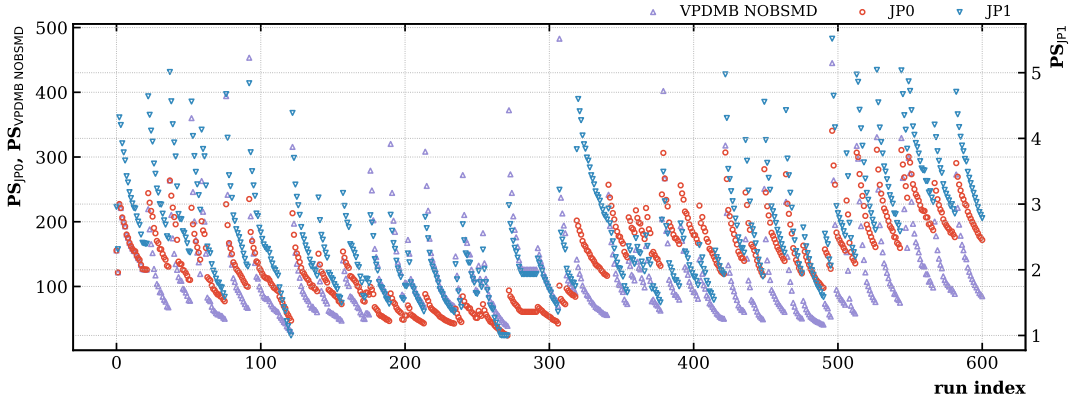


Figure 3.1: Average trigger prescale factors for the runs in the "good" dataset. The prescale factors follow trends of the collision luminosity as the beam decays during each fill.

Jet Patch Trigger	Energy threshold, DSM ADC units	Energy threshold, GeV
JP0	20	≈ 3.5 GeV
JP1	28	≈ 5.4 GeV
JP2	36	≈ 7.3 GeV

Table 3.1: Thresholds for the Jet Patch triggers that were used in the present dataset.

3.2 Trigger setup

Three jet patch triggers are used in this analysis JP0, JP1 and JP2, with thresholds set according to Table 3.1. The prescale factors were about 141 for the JP0 trigger and 2.6 for JP1, while the JP2 trigger was not prescaled at all (or, equivalently, has a fixed prescale factor of 1). The prescale factors varied as the beam luminosity decreased during a fill, as shown in Fig. 3.1.

Two other useful triggers were a "Min-bias" trigger called "VPDMB" and a "Zero-bias" trigger. The VPDMB trigger required coincidental hits in the East and West VPD and a timing difference of those hits had to correspond to an origin vertex with position $|v_z| < 30$ cm. The forward location of the VPDs allows one to sample many of the collision events without biasing too much towards large momentum transfers, hence the name "Min-bias". The Zerobias trigger was a simple trigger firing every 524289 bunch crossings, independently of any other trigger, with a fixed prescale of 10. The Zerobias trigger data would predominantly sample detector states in-between the jet events of interest, which allows the study of pile-up and other kinds of backgrounds. This has a particular application with regards to producing an "embedding simulation", which will be discussed in the next chapter.

3.3 Bunch crossings

The abort gaps in each of the two RHIC beams were aligned to overlap at the interaction point of the PHENIX experiment (at the 8 o'clock position in the RHIC rings), whereas at

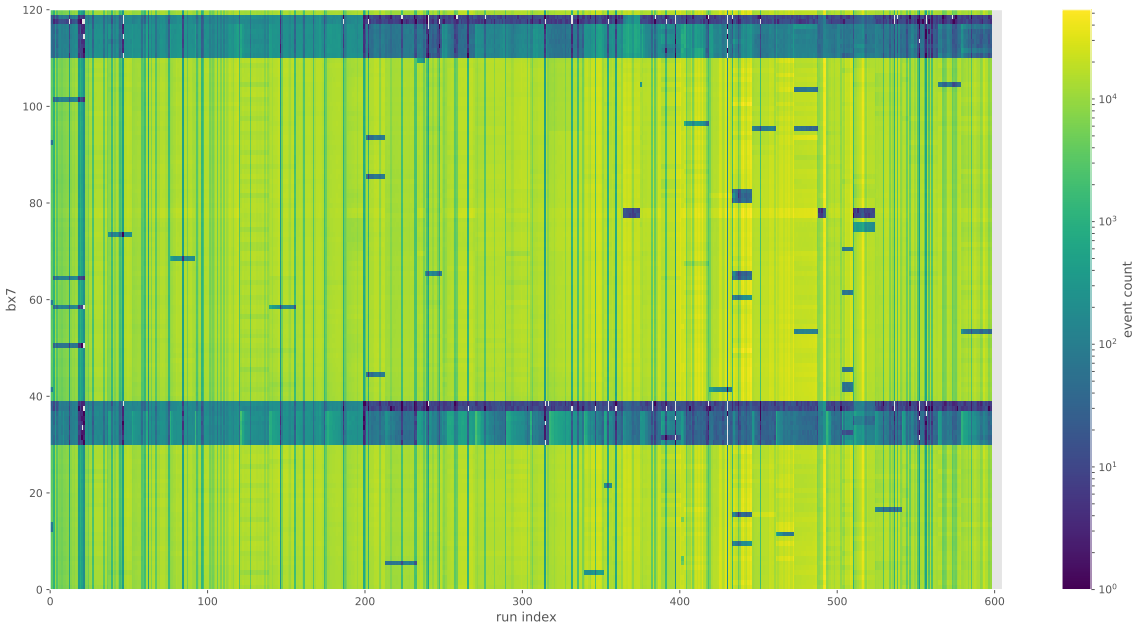


Figure 3.2: Number of events in a given run for a given bunch crossing index "bx7". The two horizontal dark bands are due to the abort gaps.

the STAR experiment the empty bunches from the gaps in one of the beams were overlapping with the filled bunches of the other beam. As seen in Fig. 3.2, this induces a background that is about four orders of magnitude smaller than the signal, and can be neglected.

3.4 Event reconstruction

The digitized signals from subdetectors are recorded online will often not readily reflect the properties of the physical particles that caused those signals. Some additional processing is required as the raw data will often need an offline calibration and, in case of track reconstruction, also need to perform computationally expensive algorithms. The event reconstruction operations are specific to each subdetector and not overly specific to any particular physics analysis.

3.4.1 Global track reconstruction

The information from the TPC coming in the form of drift times and charges sensed by the TPC pads in their respective positions needed to be converted to information about physical tracks such as their position, momentum vector and specific inionzation (energy deposition) dE/dx. The reconstruction process consisted of several steps:

1. Signals from individual pads within a single pad row that meet a certain criteria were converted to so-called TPC "hits". The hits are characterized by the ID of the pad row (layer) they belong to, their position (which is averaged from up to three pads) and the

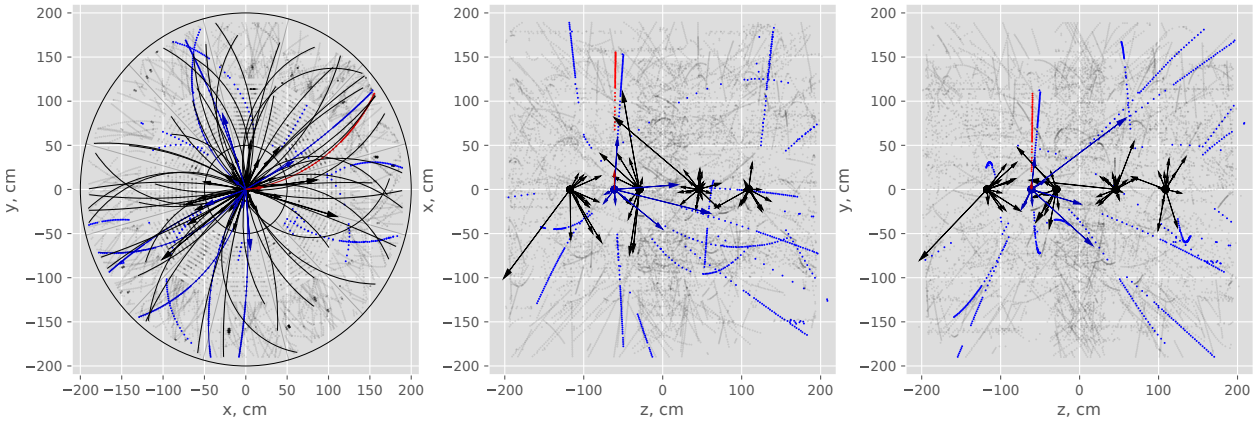


Figure 3.3: 2D projection of TPC hits, tracks (drawn as arcs on the leftmost panel and arrows on the other two panels) and primary vertices. An increased hit density in the area $|z| < 20$ cm corresponds to out-of-bunch hits being recorded (this allows to identify some of the pile-up tracks among those that appear to cross the TPC membrane).

magnitude of the detected ionization. A description of an early "hit finder" algorithm can be found in [48].

2. The positions of the hits were corrected for distortions due to the E and B field inhomogeneities, misalignments and noncollinearities[1] in those two fields. These include contribution of the time-dependent E field from the ion space charge and from ions leaking (predominantly around the gating grid) back into the TPC drift volume.[2]

3. Hits were combined into the "global tracks". This is done using a track reconstruction algorithm. The present analysis relies on the data reconstructed using the "`Sti`"[3] algorithm (ported from an algorithm based on Kalman-filtering originally developed for the ALICE experiment [53]).

3.4.2 Vertex finding

If multiple global tracks in the event are pointing to the same location (usually constrained to be along the beam line), this location is taken as a "primary vertex" seed. Then the information from the global tracks pointing to the seed, along with information from the fast detectors, was used to give vertices "rank" which rates our confidence that this vertex was a result of a beam-beam interaction occurring in the same bunch crossing as the recorded event. The parameters that contribute to the vertex ranking were based on the properties of its tracks such as the number of tracks that are matched to fast detector hits and the total p_{T} of tracks. The global tracks were then refit again using the vertex position as an additional point on the track, the momentum of the track is corrected for the measured energy loss, and

[1] Contributions to $E \times B$ create axial shifts due to the Lorentz force.

[2] See [49] and [50] for details about the correction, [51] for relevant formalism, and [52] for illustrations of the grid leak.

[3] There exists a more sophisticated "`StiCA`" algorithm

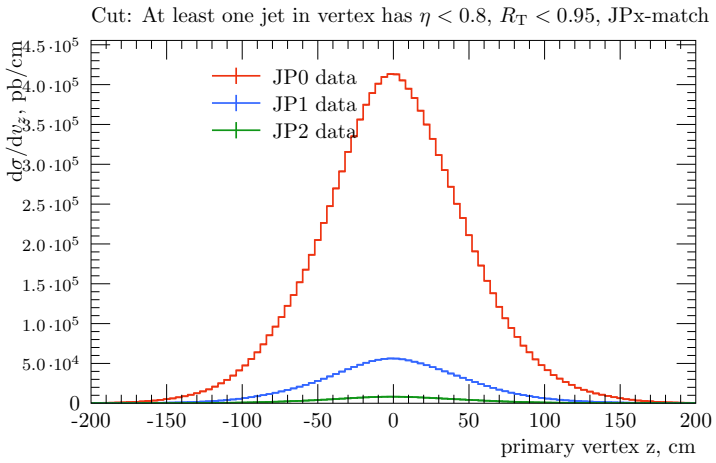


Figure 3.4: Distributions of the reconstructed vertex z coordinate for events that contain jets of interest. The distributions are dominated by the low-p_T jets.

the resulting track with slightly improved parameters is referred to as the "primary track"[4]. This dataset was reconstructed using a vertex finding algorithm called "PPV" which is the one that is usually used for the proton-proton collision data [54].

For the purposes of this analysis, each event is considered with its highest ranking primary vertex. This vertex is also required to have a ranking that is greater than zero, or the event is discarded. Additionally, there is a cut placed on its vertex z (v_z coordinate, see Fig. 3.4) coordinate to be within 60 cm from the center of the TPC. This ensures a reasonable detector acceptance and reduces uncertainties related to the unknown shape of the true vertex distribution tails. No additional cut on the vertex position is needed in the transverse plane as one is already enforced due to a beamline constraint applied in the vertex finder itself.

3.4.3 Tower reconstruction

In each event the physical state of each tower is given by the energy deposited in it:

$$E = (\text{ADC} - \text{ADC pedestal}) \cdot \text{gain} \tag{3.1}$$

The "offline" (post data collection) calibration of the BEMC was performed [42] in two steps: first, using a plethora of hits from minimum ionizing particles[5] (MIPs) to do a "relative calibration" of individual BEMC towers, and then fixing the absolute calibration with high-energy electrons.

$$C_{\text{relative}} = \frac{264 \text{ MeV}(1 + 0.056 \cdot \eta^2)}{\text{ADC}_{\text{MIP}} \cdot \sin\theta} \tag{3.2}$$

where η is the pseudorapidity of the tower, and θ is the polar angle (which can be expressed as a function of η), and ADC_{MIP} is the ADC value of the MIP peak (after subtracting

[4]Primary tracks will be referred to as "tracks" unless noted otherwise

[5]These are the charged particles that pass through the detector tower without interacting hadronically.

the pedestal) for the tower. The numerical constants were measured using beam tests and acceptance simulations [55].

The absolute calibration with electrons used the momentum p measured in the TPC:

$$C_{\text{absolute}} = \frac{C_{\text{relative}}}{\left\langle \frac{E}{p} \right\rangle} \tag{3.3}$$

The absolute calibration was performed for 40 rings of towers[6] at a specific η, and for towers read out by the same digitizer crate.

The limited acceptance of the TPC in the forward region makes it unfavorable for use in the EEMC calibration. Instead, the EEMC calibration relies on MIPs identified with its fine-grained[7] Shower Max Detector (ESMD) to deposit an expected 20 MeV for normal incidence to tower.

The EMC towers provide limited information about particle trajectories. Assuming a neutral particle with an origin somewhere along the beamline gives us the azimuthal direction φ, however the pseudorapidity η remains ambiguous. In this measurement, all available towers are interpreted in the context of the highest ranking vertex, meaning that the particle is assumed to come from the vertex to the tower location. This assumption will be true for towers of interest and false for the background towers. The contributions of the latter need to be accounted for in a simulation embedding procedure (will be discussed in Chapter 4).

For the purpose of jet reconstruction we define a "transverse energy" of the tower

$$E_T \equiv \frac{E}{\cosh(\eta)} \tag{3.4}$$

In this analysis the tower E_T is used interchangeably with p_T – it is assumed that either the energy deposited in the towers comes from a single photon, or that the pion mass can be neglected.

3.4.4 Jet reconstruction

The standard anti-k_{T} algorithm [8] was used with the initial set of proto-jets populated from lists of reconstructed TPC tracks and EMC towers. Some quality assurance selection criteria were applied to those tracks and towers to remove those coming from a background or those that are not expected to be precise. The assumption here is that only the tracks and towers of interest (which should be well reproduced in the detector simulation) will pass these selection cuts, and the decreased efficiency is to be corrected as a part of the unfolding procedure.

Track selection cuts

The summary of the cuts applied to the primary tracks is given in Table 3.2. Minimal track p_{T} is set to be above 0.2 GeV which corresponds to a track with a radius of curvature of 1.3 m, which is just enough for a particle to travel in the transverse (x-y) plane from the beamline, through half of the TPC active volume radius, to reach the first pad row of the

[6] Each ring was 1 tower wide, 120 towers in total

[7] Relative to a BEMC's Shower Max Detector – BSMD

Cut
• $p_{\mathrm{T}} > 0.2$ GeV
• $N_{\mathrm{hits}} > 12$
• Must have hits in the outer TPC [56]
• $(N_{\mathrm{hits}}/N_{\mathrm{hits\ poss.}}) > 0.51$
• $\mathrm{DCA}^{\mathrm{max}} = \begin{cases} 2\text{ cm} & \text{if } p_{\mathrm{T}} < 0.5\text{ GeV} \\ (2.5\text{ cm} - p_{\mathrm{T}} \cdot (1\text{ cm/GeV})) & \text{if } 0.5\text{ GeV} \leq p_{\mathrm{T}} < 1.5\text{ GeV} \\ 1\text{ cm} & \text{if } 1.5\text{ GeV} \leq p_{\mathrm{T}} \end{cases}$

Table 3.2: Summary of the QA cuts applied to the individual reconstructed primary tracks

outer TPC. The other cuts are put in place on the minimum number of TPC hits used to fit the track to ensure that the tracks have a reasonable momentum resolution. A cut on the ratio of the number of TPC hits to the estimated number of possible hits to be above 0.51 is imposed in order to suppress the double counting of tracks that can be accidentally reconstructed as two separate tracks.

For the high-p_{T} tracks we run into degrading momentum resolution. The approximate linear dependence $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} \sim p_{\mathrm{T}}$ extrapolates to up to $\simeq 50\%$ at $p_{\mathrm{T}} = 30$ GeV [57]. It appears that the data processing has a defect that sometimes causes physical particles with p_{T} lower than 30 GeV be reconstructed as tracks with $p_{\mathrm{T}} > 30$ GeV (see Fig. 3.5). A dramatic version of this effect was observed in a di-jet analysis for Run 9 [58] with most intense effect in the TPC sector 20. In Fig. 3.6 it is seen that the expected trend of average R_{T} (defined in Eq. (3.5)) going to $\sim 30\%$ is only restored with this cut applied. Another dramatic improvement from the cut is seen in the di-jet imbalance in Fig. 3.7. This quantity allows one to estimate the jet p_{T} resolution, which is expected to improve at high jet p_T, however the high-p_{T} tracks are seen to cause a significant imbalance.

Tower selection cuts

The selection criteria for the towers are summarized in Table 3.3. The cuts ensure that the towers are sufficiently above the noise level in various measures.

Hadronic subtraction

Charged particles are able to leave both a track in the TPC and a hit in the EMC towers. The dominant species of charged particles in jet events are charged hadrons. The calorimeter response to hadrons in the majority of cases results in a deposition of only a fraction of the energy [59] (10-17% for charged hadrons [60]), whereas the TPC measures the full momentum. In the procedure where the jets are reconstructed from both tracks and towers, including the

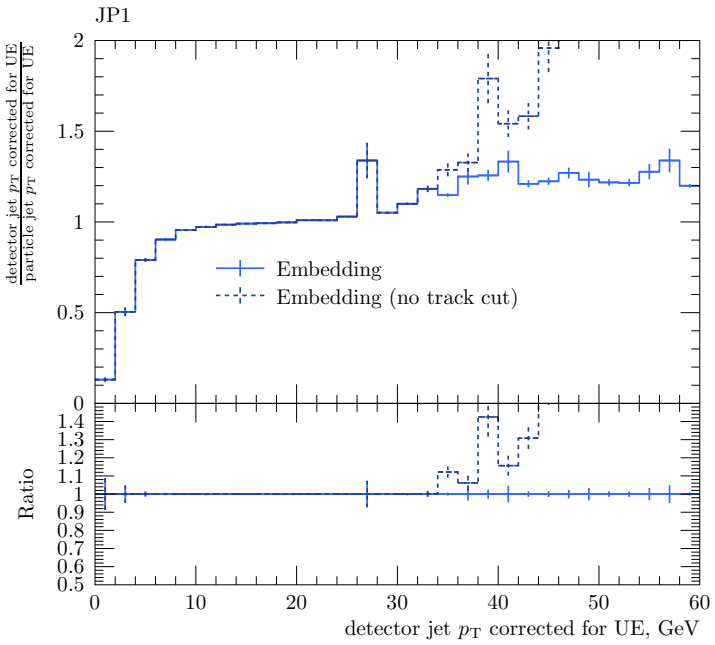


Figure 3.5: Ratio of the reconstructed jet p_{T} to the true jet p_{T} estimated from an embedding simulation for jets dominated by tracks ($0 < R_{\mathrm{T}} < 0.2$).

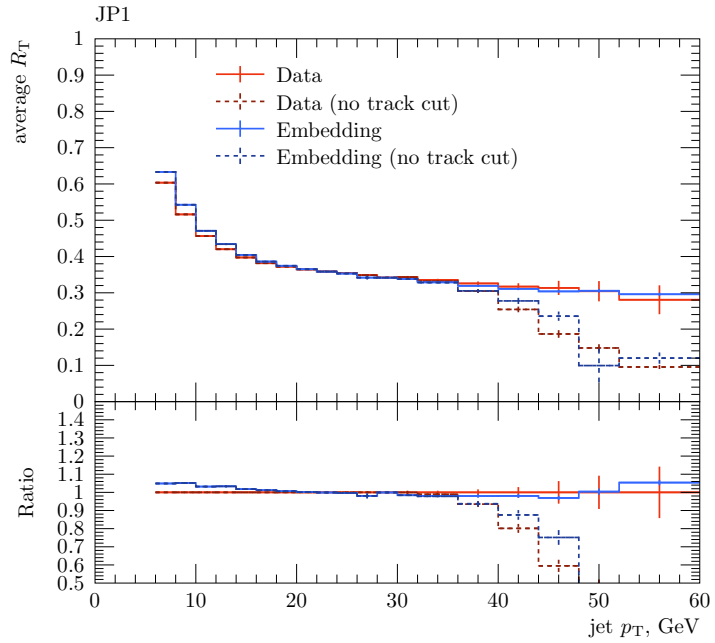


Figure 3.6: Average jet R_{T} plotted versus jet p_{T} for the JP1 data and an embedding simulation.

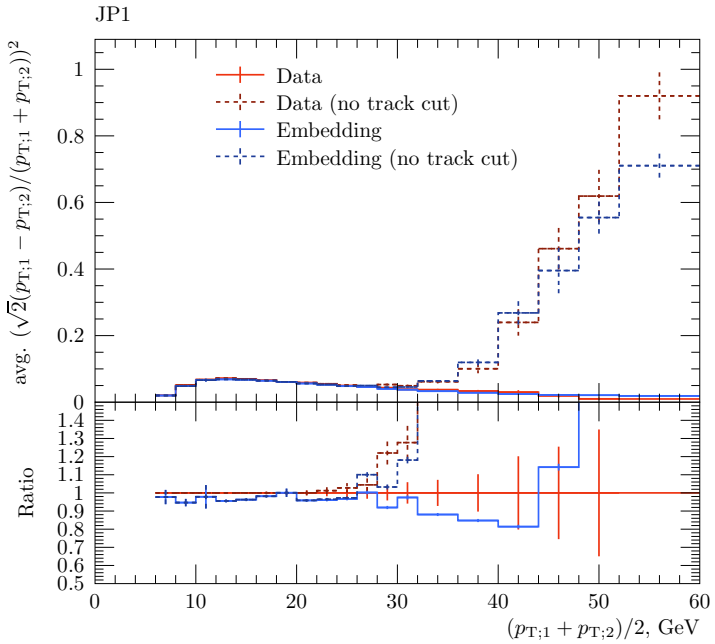


Figure 3.7: Average squared di-jet imbalance for the JP1 data and an embedding simulation.

Cut
• tower not marked as bad in the database (e.g. a hot tower)
• tower $E_T > 0.2$ GeV
• tower ADC $-$ tower pedestal > 4
• tower ADC $-$ tower pedestal $> 3 \cdot$ tower pedestal RMS

Table 3.3: Summary of the QA cuts applied to the individual reconstructed towers

Cut
• Must belong to the event's highest ranking vertex with a positive (non-pileup) rank and $\lvert v_z \rvert < 60$ cm
• $\lvert \eta_{\text{detector}} \rvert < 0.8$
• $\lvert \eta \rvert < 0.8$
• Jet must be within $\lvert \Delta\eta \rvert < 0.6$ and $\lvert \Delta\varphi \rvert < 0.6$ vicinity of the center of a Jet Patch that satisfied the trigger threshold
• $p_{\text{T}} > \begin{cases} 5 \text{ GeV} & \text{for JP0} \\ 6 \text{ GeV} & \text{for JP1} \\ 8.4 \text{ GeV} & \text{for JP2} \end{cases}$
• Jet $R_{\text{T}} < 0.95$
• Jet $R_{\text{T}} > 0$ (imposed implicitly via the Jet Patch matching requirement)
• Jet must not contain a track with $p_T > 30$ GeV

Table 3.4: Summary of the QA cuts applied to the individual reconstructed detector jets

E_T associated with the towers hit by charged hadrons leads to double counting of energy. Additionally, there is an uncertainty associated with our ability to correctly simulate our EMC detector's response to those hits (to be discussed in Section 6.5.2). Hence, a procedure is implemented to correct the towers which have primary tracks pointing to them. The energy of the tower is reduced by the momentum of the pointing track, and if the latter is greater, then the tower is removed completely. Only the primary tracks originating at the highest ranking primary vertex are used, they also have to satisfy the quality assurance cuts from Table 3.2.

Jet selection cuts

The jets reconstructed from tracks and towers are subject to selection cuts summarized in Table 3.4. For each jet a "detector pseudorapidity" η_{detector} is calculated as pseudorapidity of the vector from the beamline at $z = 0$ to the intersection of the jet thrust axis with the BEMC[8]. A limit on η_{detector} ensures that the jet, regardless of the vertex position, points within a region of appreciable BEMC and TPC acceptance. The physical jet rapidity (a rapidity of its thrust axis) is also limited to match a cut that will be imposed on particle jets in the definition of the cross section. The jets are also required to point close to the center of a jet patch that was above the trigger threshold. An important quantity of "transverse

[8]Assumed to be a cylinder with $R = 225$ cm

neutral energy fraction" R_{T} is defined for a detector jet as:

$$R_{\mathrm{T}} \equiv \frac{\sum p_{T;\mathrm{tower}}}{\sum p_{T;\mathrm{track}} + \sum p_{T;\mathrm{tower}}} \tag{3.5}$$

We are generally looking for jets that are simultaneously observed by both the TPC and EMCs to avoid various backgrounds. Hence a cut on R_{T} is placed to avoid the extremes of track-dominated or tower-dominated jets.

3.5 Quality assurance

3.5.1 L0 trigger

The radiation present during RHIC operations can cause bit flips in the BEMC electronics, which may result in excessive jet patch triggers fired, either by creating hot towers or by directly affecting the trigger electronics. The latter case can be partially rectified by reevaluating the trigger states in post-processing by software called a "trigger simulator" that implements the DSM board algorithms to operate on recorded tower ADC values. The hot towers can also be discovered individually, and by masking them out before applying the trigger simulator, we can filter out the extra triggered events again.

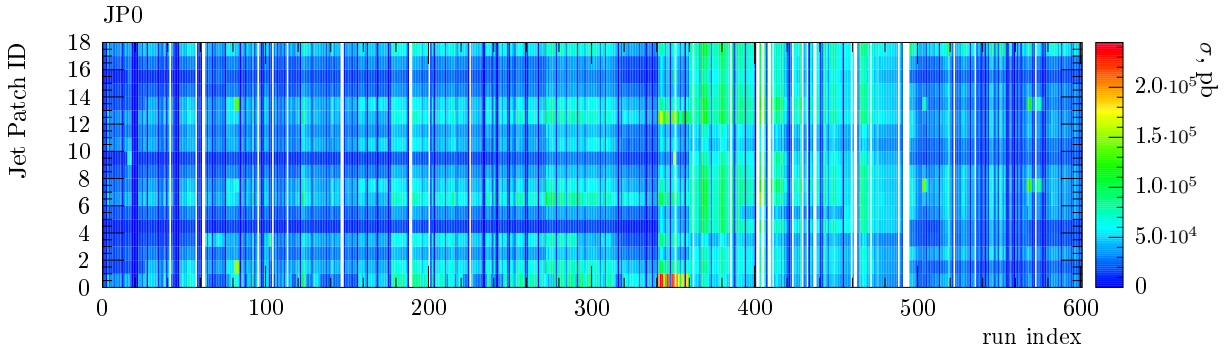
Figure 3.8 shows how the simulated triggers were distributed among different jet patches for different thresholds. Some hot patches remain visible. The JP0 and JP1 trigger rates have a change in behavior around runs with indices 360 and 491. The nature of the anomaly is not understood, however it is reproduced in the simulation.

3.5.2 The jet R_{T} for mono-jet and di-jet events

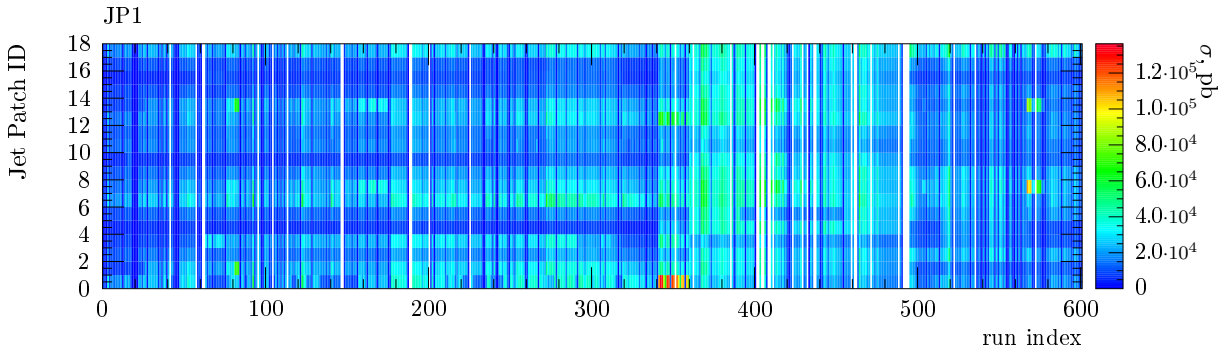
In previous years of STAR operation there was an excess of background jet triggers caused by the RHIC beams blasting the calorimeters due to insufficient shielding in the tunnels. A way to study such occurrences is to compare the jet R_{T} distributions for events containing only a single jet and events with two jets [61]. Background coming from beam sprays would predominantly show up as mono-jet events with particles coming not from a valid vertex, so there would be few or none reconstructed TPC tracks in the jet, only EMC towers, leading to $R_{\mathrm{T}} \simeq 1$. At the same time, selection of the events with di-jets would be mostly satisfied by the back-to-back jets from the physical scattering events. As seen in Fig. 3.10, there is no discrepancy at high R_{T} values between the shapes of the R_{T} distributions for mono-jet and di-jet events. The absence of the background is not surprising given that lead/iron shielding walls were installed in the RHIC tunnel before year 2012 to avoid this problem.

3.5.3 Track DCA and Sector 20

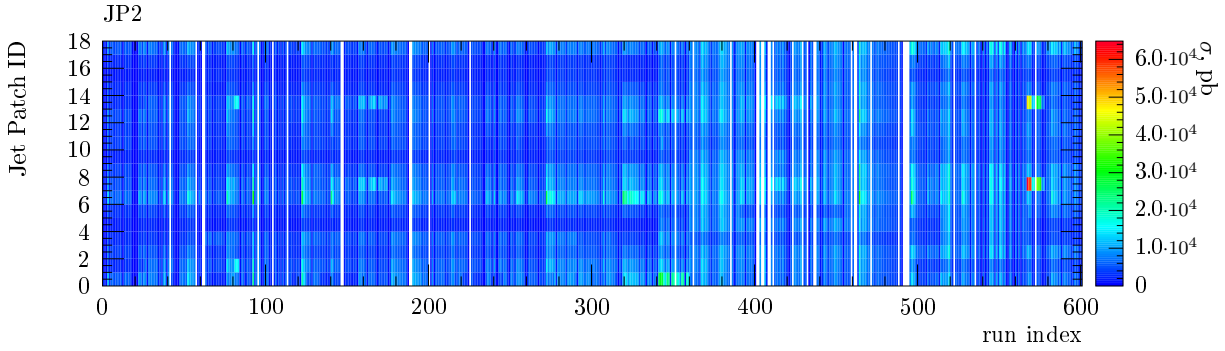
There are different calibrations [62] that need to be applied to the raw TPC data: timing "T0" and slewing corrections, drift velocities, static electric and magnetic fields, and detector alignment. A "Space Charge" effect is caused by density of residual ions varying along the radial coordinate in the TPC volume. This creates a radial component of the electric field



(a) JP0

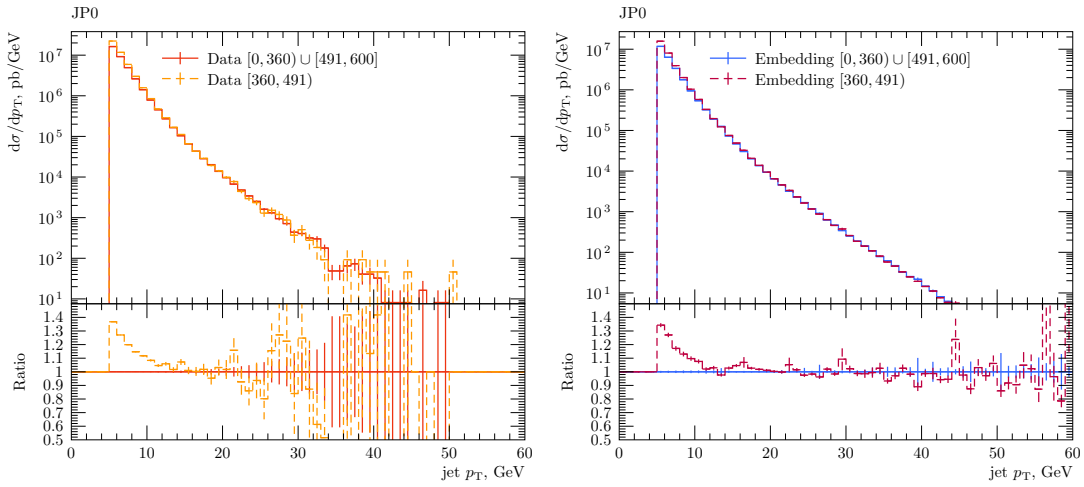


(b) JP1



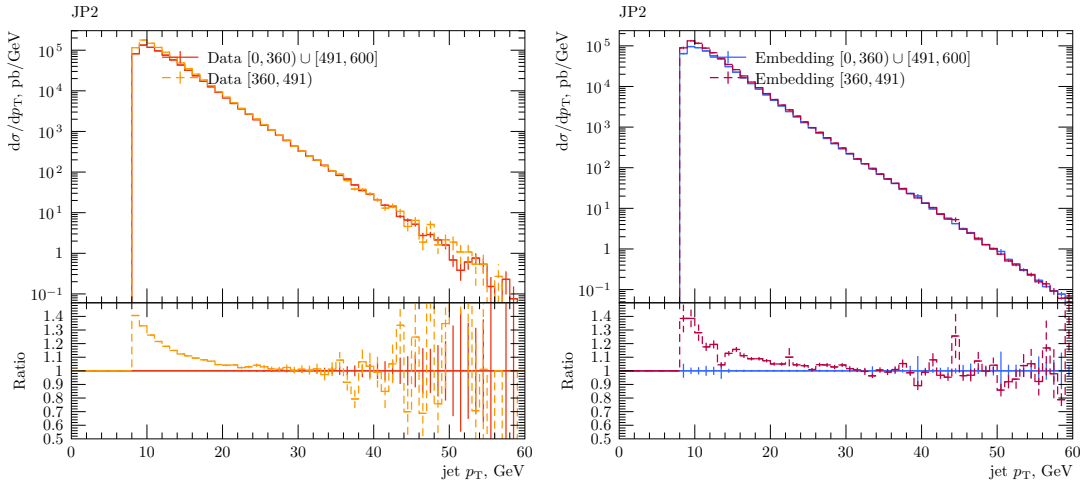
(c) JP2

Figure 3.8: Number of events in the data for a given run with a given BEMC jet patch with ADC value (simulated offline) above the trigger threshold.



(a) JP0 data

(b) JP0 embedding simulation



(c) JP2 data

(d) JP2 embedding simulation

Figure 3.9: Jet p_{T} spectra for the JP0 trigger in runs within and outside the outlier region.

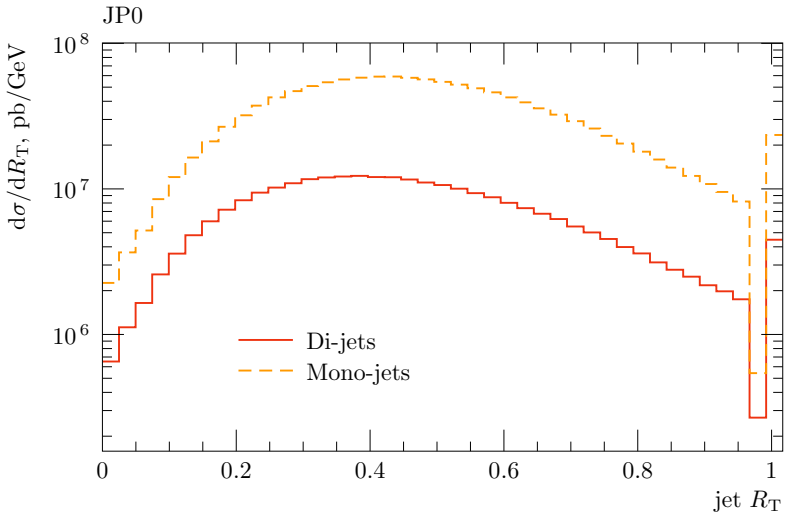


Figure 3.10: The R_T distribution for jets in mono-jet and di-jet events for the JP0 trigger.

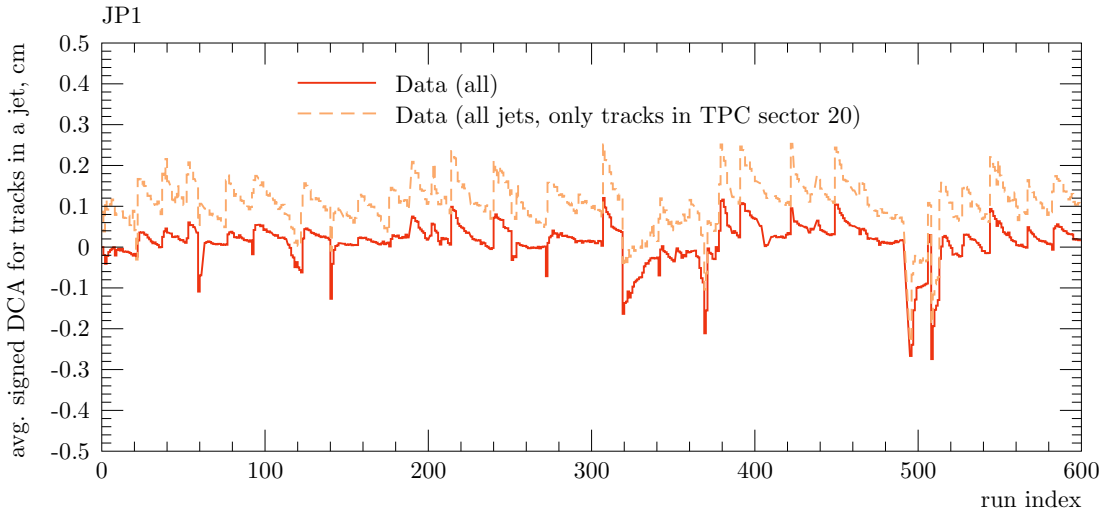


Figure 3.11: Average value of the signed DCA in a given run for the JP1 trigger. The graphs of values for other JP triggers and VPDMB look similarly.

contributing a non-zero component to $\vec{E} \times \vec{B}$ which causes an additional distortion in the process of the drift along the z axis. This distortion affects the transverse momentum of the reconstructed track. It can vary on a time scale of seconds and is dependent on the instantaneous collider luminosity, but may also depend on other conditions that are difficult to monitor. Fortunately, the deviations can be detected by looking at the signed Distance of Closest Approach (signed DCA) of the global tracks that are associated to the primary tracks from the analysis. Looking at it in Fig. 3.11, a large cusp was observed in the run with index 491. This, and the three following runs, were excluded from the analysis to avoid the vertexing efficiency variations that cannot be reproduced in the simulation.

Special attention is paid to the tracks measured in sector 20 of the TPC. Before the "iTPC" upgrade made in 2018 there was a hardware issue causing the inner part of TPC sector 20 to have a low efficiency. That created complications for performing the Space Charge calibration for this particular sector. As can be seen in Fig. 3.11, this sector has a positive signed DCA offset. The offset, on the order of 0.1 cm, is tolerable as it is well below the DCA cut (see Table 3.2) and smaller than the width of the DCA distribution. This was also confirmed by removing jets that contain tracks in sector 20 and observing no unusual change in the basic distributions.

Chapter 4

Detector simulation

Data alone provide the sample of jets reconstructed from the tracks and towers measured in the detector; these are called jets at the "detector level" or, simply, "detector jets". In order to facilitate comparisons with theory and between experiments, a jet cross section that is defined in a way that is free from the specific details of the STAR detector would be preferred. Practically, that means it has to be defined in terms of jets comprised of the physical particles, at a "particle level", referred to as "particle jets". In order to get to that, one needs to be able to estimate, for a given detector, how the transverse momentum of the particle jets is related to the transverse momentum of the detector jets.

The effects that need to be considered include the acceptance of the detectors, as well as, more generally, their efficiencies, their resolutions, and their backgrounds. This information can be extracted from the full event computer simulations. They allow, knowing the species of all the outgoing primary particles and their initial position and momenta, to predict a possible measured event as it would be seen by the detector. In the case of STAR, a simulation framework based on GEANT 3 [63] propagates the energy deposition along paths traveled by the primary particles and their secondary particles produced as the result of particle decays and interaction with the material of the detectors. Then the resulting hits and energy depositions can be converted to their associated digital values using models for signal detection technologies and electronics used in each specific detector. A special step is taken in case of the TPC to simulate the drift of the ionization clusters towards the anodes, while trying to account for at least some of the distortions. In order to simulate effects of the pile-up background, the simulated detector responses are mixed with the zero bias data in a procedure called "embedding" (to be described in Section 4.2). Finally, the resulting simulated event can be reconstructed just as was done for the data.

This chapter will discuss the details behind the production and evaluation of the simulated data sample. The details of the procedure used for relating the measured detector level cross section to a particle level cross section will be discussed in a later chapter.

4.1 Particle level simulation

As was previously mentioned, the detector event simulation requires initial events defined at the particle level. In this analysis, the Pythia 6 [5] Monte Carlo (MC) event generator is

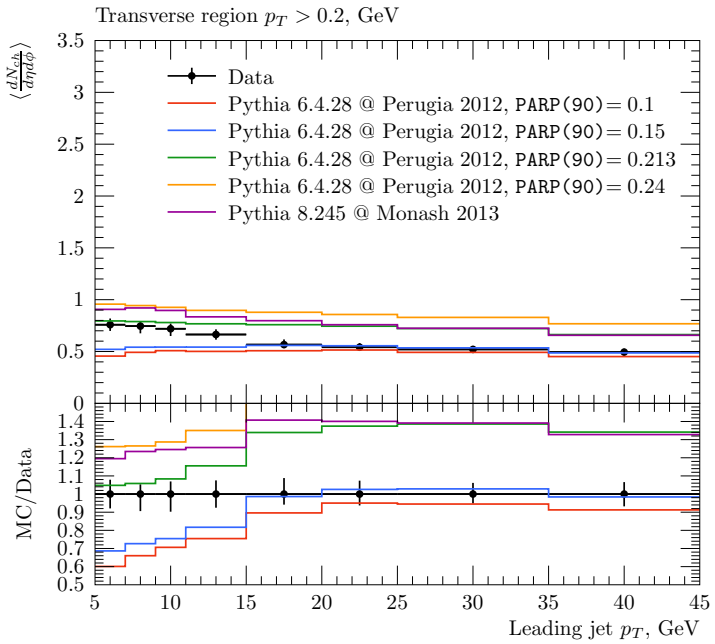


Figure 4.1: Underlying event activity measured at STAR [44] plotted as a function of the event's leading jet p_{T}. Predictions from PYTHIA with different p_{T0} tunes are shown for comparison.

used to produce such physical events. The main requirement for the generator here is for it to produce proton-proton collision events that are comparable to what we see in the data. Direct comparisons to the data are only possible after the detector simulation, which creates an ambiguity with interpreting possible discrepancies with the data. Any difference can be attributed to a possible shortcoming in the event generator model or a problem with the detector simulation.

4.1.1 Pythia Tune

The PYTHIA 6 MC generator comes with many tunable parameters, however most combinations of the settings are not expected to well reproduce the processes occurring in nature. Instead, sets of parameters called "tunes" that are known to work and reproduce various experiment's data are provided. The simulation to be used here is based on the Perugia 2012 tune [64] that uses the CTEQ6 PDF set [65].

PARP(90)

The PYTHIA tunes from its standard set were primarily developed to describe the data from the Tevatron and the LHC, and they are not always able to describe the data at RHIC energies using default parameter values of the model. A modification to the Perugia 2012 tune was developed at STAR [42] to best fit the charged pion cross section measured at STAR at $\sqrt{s} = 200$ GeV [66, 67] to remove a discrepancy for the pion production with a transverse momentum in the 0-3 GeV range. This was achieved by modifying the value of a

PARP(90)	$p_{\mathrm{T0}}(\sqrt{s} = 200\ \mathrm{GeV})$
0.24	1.13 GeV
0.213	1.24 GeV
0.19	1.35 GeV
0.15	1.55 GeV
0.1	1.86 GeV

Table 4.1: Correspondence between the PARP(90) values and p_{T0} for Perugia 2012 at $\sqrt{s} =$ 200 GeV.

p_{T0} parameter that governs the handling of the Multiple Parton Interaction (MPI), which was implemented by setting the value of the PARP(90) parameter to 0.213.

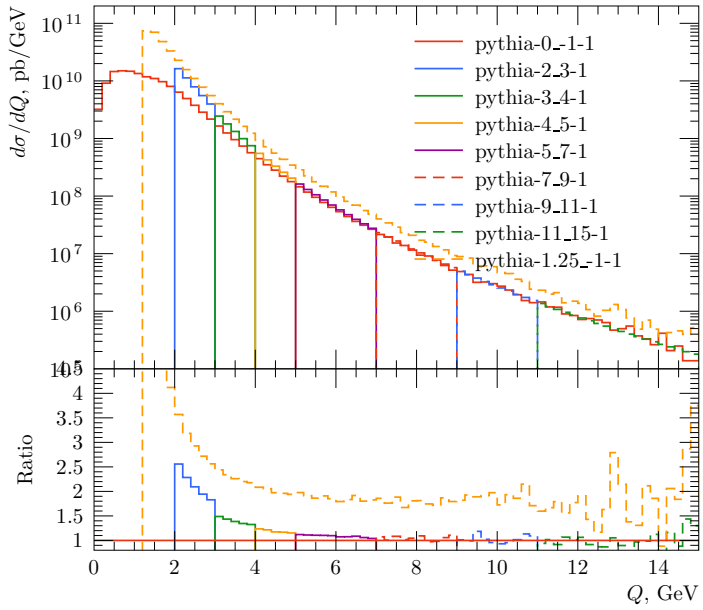
This tune adjustment improves the agreement with the measurement of the Underlying Event activity done at STAR [44]. Figure 4.1 shows the UE measurement in comparison to different PARP(90)-modifications of the Perugia 2012 tune for PYTHIA 6. The correspondence of those PARP(90) values to the regularization parameter p_{T0} is given in Table 4.1. In principle, the UE tune can be improved, but this was not done in order to stay consistent with another inclusive jet cross section measurement from Run 12 at STAR from a different collision energy of $\sqrt{s} = 510$ GeV [68].

4.1.2 Partonic p_{T} bins

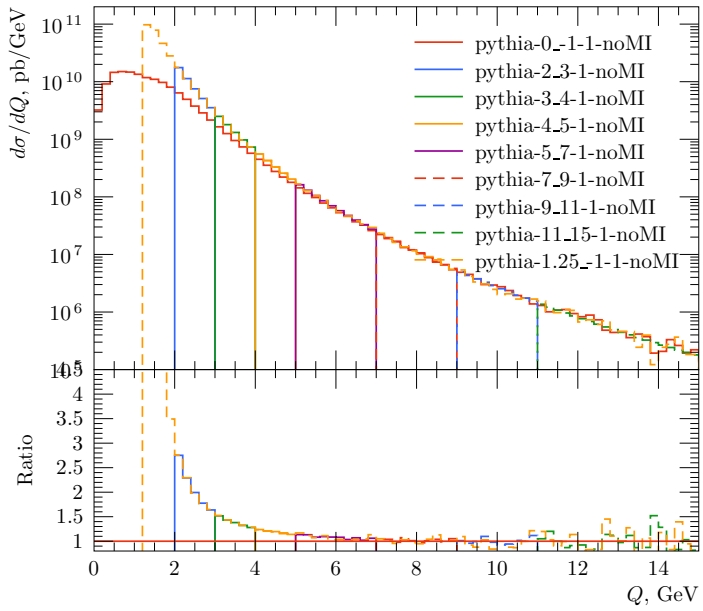
The particle level events required for simulation have to span a wide range of phase space and provide enough statistical precision for jets with p_{T} up to ≈ 50 GeV. This presents a challenge for an event generator that produces unweighted events, as the large difference in the magnitude of the differential jet cross section at low and at high jet p_{T} makes the high-p_{T} events always constitute a tiny fraction of a sample, requiring a large simulation sample (with a large cost in terms of the CPU time).

The solution to this is to do the generation with events of unequal weights. In STAR this is commonly done by slicing the phase space by the value of the transverse momentum transfer $\hat{p}_{\mathrm{T}}$ characteristic of the parton-parton scattering process. Each bin of $\hat{p}_{\mathrm{T}}$ has a fixed physical total cross section $\sigma_{\hat{p}_{\mathrm{T}}\ \mathrm{bin}}$, but an arbitrary number of events $N_{\hat{p}_{\mathrm{T}}\ \mathrm{bin}}$ can be generated within that bin using low-level cuts in PYTHIA. A set of adjacent $\hat{p}_{\mathrm{T}}$ bins with different inverse luminosity weight factors $\sigma_{\hat{p}_{\mathrm{T}}\ \mathrm{bin}}/N_{\hat{p}_{\mathrm{T}}\ \mathrm{bin}}$ will then represent a complete physical sample of events.

The bin boundaries were chosen to cover an interval of $\hat{p}_{\mathrm{T}}$ from 2 GeV to $+\infty$. Events with $\hat{p}_{\mathrm{T}}$ less than 2 GeV are not considered as they do not contribute a significant number of jets to the analyzed sample, which requires jets that have a larger transverse momentum of $p_{\mathrm{T}}^{\mathrm{jet}} \gtrsim 5$ GeV. The highest lower-edge of a bin is 55 GeV, which is needed to cover detector jet p_{T} of ~ 50 GeV. The list of bins and their parameters is given in Table 4.2.



(a) With MPI enabled



(b) With MPI disabled

Figure 4.2: Ratio of samples generated with PYTHIA 6 for $\hat{p}_{\mathrm{T}}$ cuts $[1.25, \infty)$ $[2, 3)$, $[3, 4)$, $[4, 5)$, etc., to a $[0, \infty)$ reference sample. Only the version with the MPI enabled (on plot a) shows discontinuities that bin weights are meant to address.

$\hat{p}_{\mathrm{T}}$ bin, GeV/c	$N_{\hat{p}_{\mathrm{T}}\text{ bin}}$	$\sigma_{\hat{p}_{\mathrm{T}}\text{ bin}}$, pb	$f_{\text{partonic. bin}}$
2 – 3	3 588 051	$9.00 \cdot 10^9$	1/1.228
3 – 4	3 588 051	$1.46 \cdot 10^9$	1/1.051
4 – 5	3 587 534	$3.54 \cdot 10^8$	1/1.014
5 – 7	3 588 051	$1.51 \cdot 10^8$	
7 – 9	3 588 051	$2.49 \cdot 10^7$	
9 – 11	3 588 051	$5.84 \cdot 10^6$	
11 – 15	3 588 051	$2.30 \cdot 10^6$	
15 – 20	3 587 313	$3.42 \cdot 10^5$	
20 – 25	3 587 520	$4.57 \cdot 10^4$	
25 – 35	2 391 292	$9.72 \cdot 10^3$	
35 – 45	2 390 073	$4.69 \cdot 10^2$	
45 – 55	1 195 609	$2.69 \cdot 10^1$	
55 – ∞	1 195 609	$1.43 \cdot 10^0$	

Table 4.2: List of $\hat{p}_{\mathrm{T}}$ bins and their respective parameters.

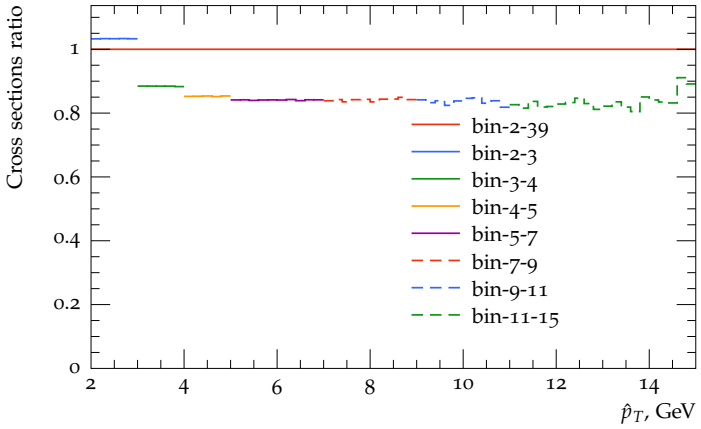


Figure 4.3: Ratio of samples generated with PYTHIA 6 for $\hat{p}_{\mathrm{T}}$ cuts $[2, 3)$, $[3, 4)$, $[4, 5)$, etc., to a $[2, 40)$ reference sample.

4.1.3 Partonic p_{T} bin weights

Individual partonic p_{T} samples are generated using Pythia's settings for minimum and maximum partonic p_{T} cuts defined by the `CKIN(3)` and `CKIN(4)` variables. Using these values has a few side effects when the minimum cutoff value is above the regularization scale `PARP(82)`. One such effect comes from the treatment of the MPI in Pythia that makes it report a bin cross section that is slightly too large [5]. This results in discontinuities in the differential (in partonic p_{T}) cross section at the boundaries between different partonic bin samples whenever the MPI is enabled (see Fig. 4.2). This effect is corrected in post-processing by multiplying each bin's cross section by a bin weight factor $f_{\text{partonic. bin}}$ which has the physical meaning of an inverse of a probability for generated hard events in a specific partonic p_{T} bin to be harder than generated MPI events.

To determine the bin weight, one relies on the fact that they tend to approach 1 for `CKIN(3)` $>\sim 5$ GeV [69]. The partonic cross section distribution was generated in bin $[2, \infty)$ (`CKIN(3)=2`, `CKIN(4)=-1`) and in bin $[11, 15]$ where $f^{[11,15]}_{\text{partonic. bin}}$ is assumed to be equal to 1. The ratio of these two distributions in the region of their overlap is then constant and equal to $f^{[2,\infty)}_{\text{partonic. bin}}$. Then the factor of any other bin (say $[2, 3]$) can be determined as the ratio of a corrected $[2, \infty)$ distribution to the distribution in question (see Fig. 4.3). The resulting values of bin weights are given in Table 4.2. This procedure can be summarized with the following equations:

$$f^{[\hat{p}_{\mathrm{T}}^{\min},\hat{p}_{\mathrm{T}}^{\max}]}_{\text{partonic. bin}} = \frac{f^{[2,\infty]}_{\text{partonic. bin}} \cdot \int_{\hat{p}_{\mathrm{T}}^{\min}}^{\hat{p}_{\mathrm{T}}^{\max}} \mathrm{d}\sigma^{[\hat{p}_{\mathrm{T}}^{\min},\hat{p}_{\mathrm{T}}^{\max}]}(\hat{p}_{\mathrm{T}})}{\int_{\hat{p}_{\mathrm{T}}^{\min}}^{\hat{p}_{\mathrm{T}}^{\max}} \mathrm{d}\sigma^{[2,\infty]}(\hat{p}_{\mathrm{T}})} \tag{4.1}$$

Where $\sigma^{[\hat{p}_{\mathrm{T}}^{\min},\hat{p}_{\mathrm{T}}^{\max}]}(\hat{p}_{\mathrm{T}})$ denotes the differential partonic cross section for the partonic p_{T} bin $[\hat{p}_{\mathrm{T}}^{\min}, \hat{p}_{\mathrm{T}}^{\max}]$. One can then use the $f^{[11,15]}_{\text{partonic. bin}} = 1$ assumption to calculate the value of $f^{[2,\infty]}_{\text{partonic. bin}}$:

$$f^{[2,\infty]}_{\text{partonic. bin}} = \frac{\int_{\hat{p}_{\mathrm{T}}^{\min}}^{\hat{p}_{\mathrm{T}}^{\max}} \mathrm{d}\sigma^{[2,\infty]}(\hat{p}_{\mathrm{T}})}{\int_{\hat{p}_{\mathrm{T}}^{\min}}^{\hat{p}_{\mathrm{T}}^{\max}} \mathrm{d}\sigma^{[11,15]}(\hat{p}_{\mathrm{T}})} \tag{4.2}$$

4.1.4 Soft reweighting

Another side effect of setting `CKIN(3)` in Pythia 6 to a value that is larger than the MPI p_{T0} scale ($\simeq 1$ GeV) is that it makes Pythia disable soft QCD behavior (sec. 8.9.3 of [5]). The soft QCD process, among other things, implements a different kind of regularization that is supposed to prevent the total hard QCD process cross section from exceeding the total inelastic non-diffractive cross section. The difference between the soft QCD-enabled and the soft QCD-disabled samples can be practically seen in a difference between the shapes of $\mathrm{d}\sigma/\mathrm{d}\hat{p}_{\mathrm{T}}$ generated with a $[0, \infty)$ and a $[2, \infty)$ $\hat{p}_{\mathrm{T}}$ cuts. There is no other effect of this visible in distributions of jets other than the p_{T} one.

It is not clear if the soft QCD process would be preferable over any other kind of model. However, in any case, it would be preferred to be able to evaluate that model in case it allows one to better reproduce the shape of the jet p_{T} spectrum at low-p_{T}. To recover the behavior

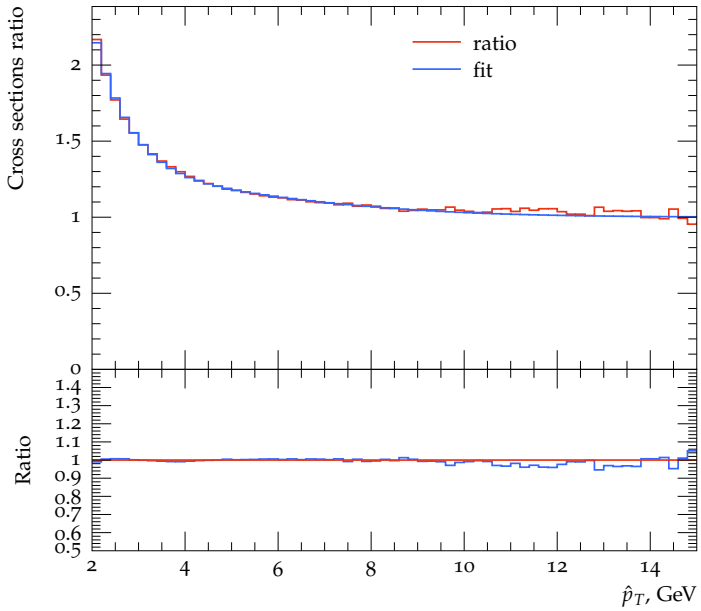


Figure 4.4: Plot of $[2, \infty)$ sample using the `CKIN(3)=2` cut and compensated for its bin weight to a $[0, \infty)$ sample. The fit line is given by $1/\omega(\hat{p}_\mathrm{T})$ (see Eq. (4.4)).

in the produced simulation sample that relies on the `CKIN(3)` cut, the event weights were corrected by giving them an additional $\hat{p}_\mathrm{T}$-dependent weight:

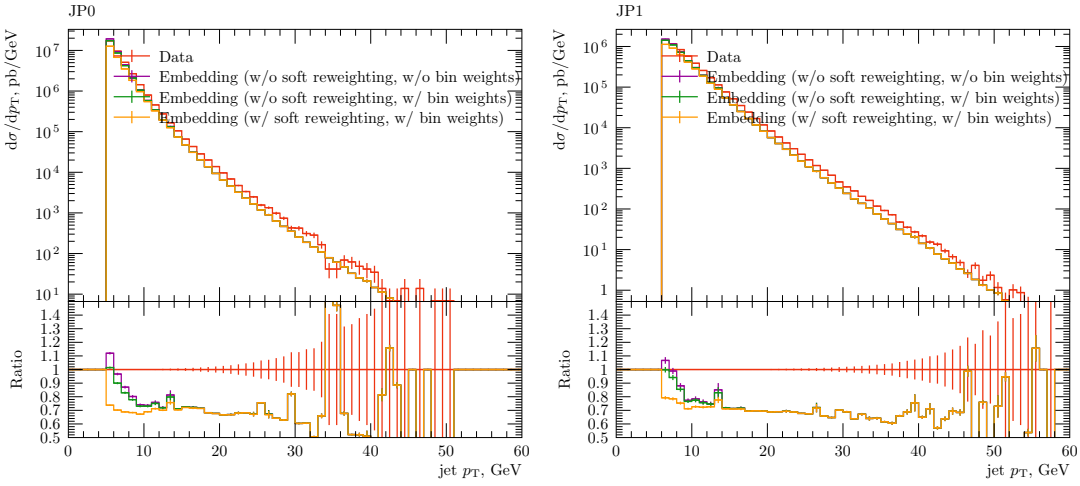
$$\omega(\hat{p}_\mathrm{T}) \equiv \frac{\sigma^{[0,\infty]}(\hat{p}_\mathrm{T})}{\sigma^{[2,\infty]}(\hat{p}_\mathrm{T})} \tag{4.3}$$

In practice, this was approximated with the following function:

$$\omega(\hat{p}_\mathrm{T}) = \frac{1}{1 + (1.22 - 0.33\,\kappa + 0.17\,\kappa^2)\exp(-0.82\,\kappa)}, \quad \text{where } \kappa = \frac{\hat{p}_\mathrm{T} - 2\ \mathrm{GeV}}{1\ \mathrm{GeV}} \tag{4.4}$$

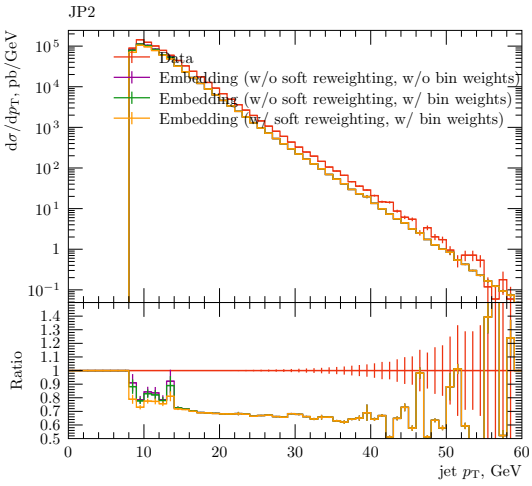
In the above equation, the 2 GeV offset was taken from the low $\hat{p}_\mathrm{T}$-bin boundary, and the coefficients in the denominator were determined by fitting the ratio of the partonic cross sections generated in the $[0, \infty)$ and a $[2, \infty)$ bins for Perugia 2012 and $\sqrt{s} = 200$ GeV (fit shown in Fig. 4.4). The effect of this reweighting on the reconstructed jet p_T spectra can be seen in Fig. 4.5.

An alternative approach to this "soft reweighting" could be to manually implement the minimal threshold on $\hat{p}_\mathrm{T}$ by discarding the generated events that fall outside of the bounds while keeping `CKIN(3)` at 0. This would be only needed for bins up to $\hat{p}_\mathrm{T} \approx 10$ GeV; the higher bins could use `CKIN(3)` and `CKIN(4)` as usual. This was done for supplementary Pythia studies in this thesis, but for the embedding simulation the soft reweighting as described above was used.

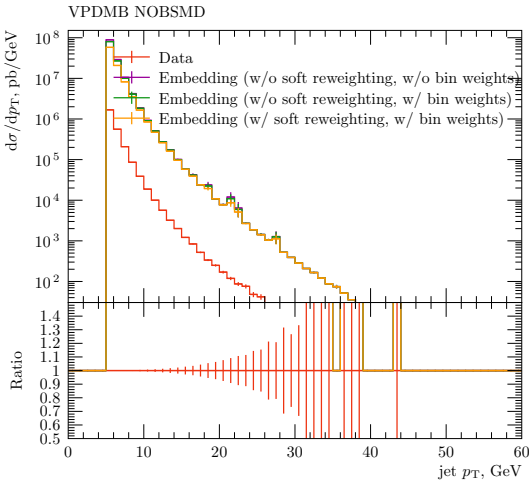


(a) JP0

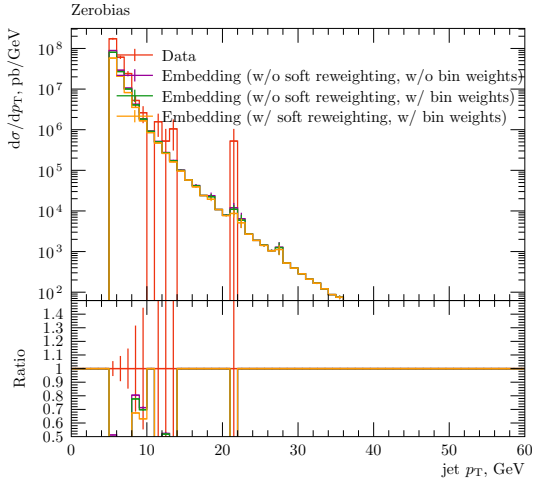
(b) JP1



(c) JP2



(d) VPDMB NOBSMD



(e) Zerobias

Figure 4.5: Effect of the soft reweighting on the detector level jet cross sections for various triggers.

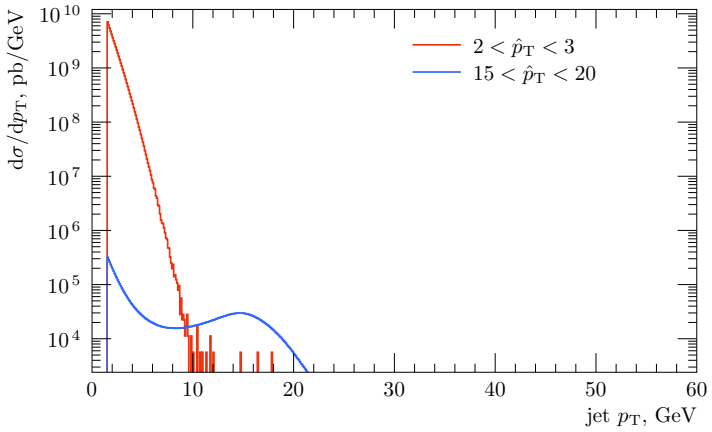


Figure 4.6: Contributions from partonic p_T bins $[2, 3]$ GeV and $[15, 20]$ GeV to the inclusive jet cross section as a function of the jet p_T.

4.1.5 Bin outliers

Having a trend of statistics that falls more slowly than the trend of the cross section, like in Table 4.2, leads to highly uneven event weight factors. This is not visible in the distribution of $\hat{p}_T$ itself, but can readily appear in the distribution of the jet p_T whenever an event with a higher weight from a lower $\hat{p}_T$ bin shoots out into a region of phase space that is otherwise populated by events with a lower weight from a higher $\hat{p}_T$ bin. The resulting effect of this is the presence of tall peaks in counts and variance in finely-binned histograms. This situation can occur through just statistical fluctuations, however it is aggravated by PYTHIA's ability to sometimes produce events with jets p_T much higher than its $\hat{p}_T$, which seemingly happens due to the MPI handling. An example of such behavior is shown in Fig. 4.6. This can not be rectified by generating extra statistics, only by removing those events.

4.2 Embedding

Straightforward detector simulation of individual PYTHIA events is not going to take into account background effects such as:

- Multiple proton-proton scatterings in the same bunch crossing (not to be confused with Multiple Parton Interactions)
- Tracks left over and tracks added during the readout in the TPC from out-of-bunch interactions (pile-up)
- Radiation from the collider operation
- TPC electronics noise and EMC tower pedestals
- Cosmic rays

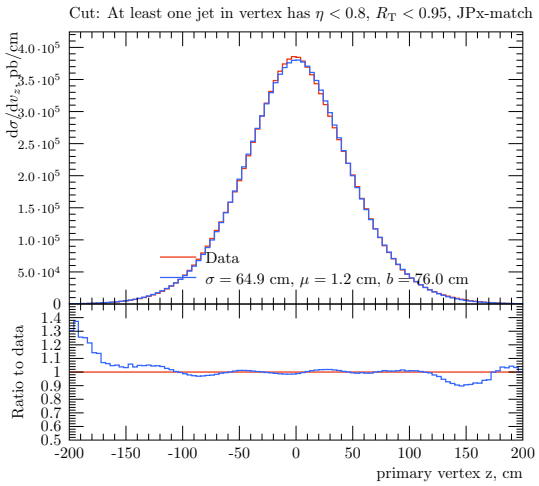


Figure 4.7: Vertex z distribution observed in the data compared to a Gaussian fit performed in the $|v_z| < 60$ cm region.

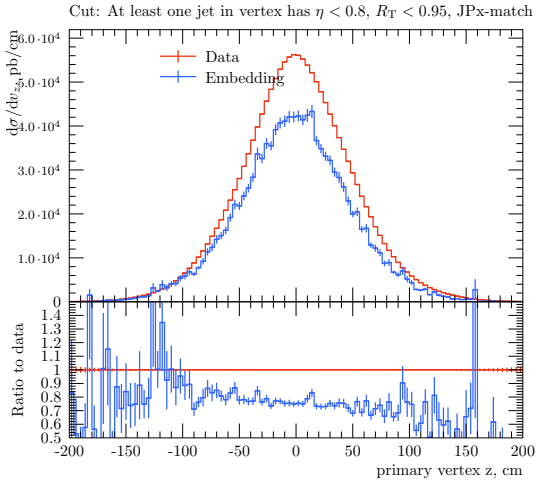


Figure 4.8: The v_z distributions observed in the data and the embedding simulation.

All of these can be taken into account by employing an "Embedding" procedure, in which the raw simulated detector responses are mixed event-by-event with the actual detector responses from events from the Zerobias trigger sample before doing the reconstruction. The Zerobias trigger fires independently of any detector condition, so it samples random detector states in random bunch crossings. This sample should effectively reproduce the possible detector states in the bunch crossings preceding and following the one that is simulated.

4.2.1 Vertex distribution

Beam-beam interactions in the data happen in a wide range of positions in the region called the "interaction diamond" for its general shape. The size and shape of the interaction vertex distribution along the z-axis depends on various parameters such as the length of the colliding bunches, the value of β^* and the crossing angle. The distribution of v_z observed

in the data is subject to possible detector biases due to the limited acceptance for vertex reconstruction and triggering. The vertexing acceptance can be assumed to be reasonably flat in the intermediate region $|v_z| < 60$ cm, which allows one to at least generally see the shape of the true distribution. A Gaussian shape is assumed to describe the data in the intermediate region and to extrapolate outside of it, as shown by the fit in Fig. 4.7. A conservative estimate for the Gaussian width of $\sigma_z = 45$ cm was used in the generation of the embedding sample. If needed, small corrections can be then performed on the simulation sample using reweighting, as described below.

Vertex reweighting

The vertex distribution in the data has long tails that are not taken into account in by the placeholder Gaussian distribution used in embedding. A reweighting to a wider Gaussian with $\sigma_{z;\text{target}} \approx 70$ cm and a hourglass effect tail suppression as discussed in Appendix A.2.1 with $b_{\text{target}} \approx 80$ cm is performed using a following factor:

$$\omega_{\text{vertex}}(v_z) = \mathcal{N} \exp\left(\left(\frac{1}{\sigma^2_{z;\text{nominal}}} - \frac{1}{\sigma^2_{z;\text{target}}}\right)\frac{v_z^2}{2}\right)\frac{1}{1+\left(\frac{v_z}{b_{\text{target}}}\right)^2} \tag{4.5}$$

where v_z is the position of the thrown vertex[1], $\sigma_{z;\text{nominal}}$ is the distribution width that was used to generate the embedding sample, and the normalization constant $\mathcal{N}$ is given by:

$$\mathcal{N} \equiv \frac{\sqrt{2\pi}\sigma_{z;\text{nominal}}}{\pi b \exp\left(\frac{b^2_{\text{target}}}{2\sigma^2_{z;\text{target}}}\right)\left(1-\text{erf}\left(\frac{b_{\text{target}}}{\sqrt{2}\sigma_{z;\text{target}}}\right)\right)} \tag{4.6}$$

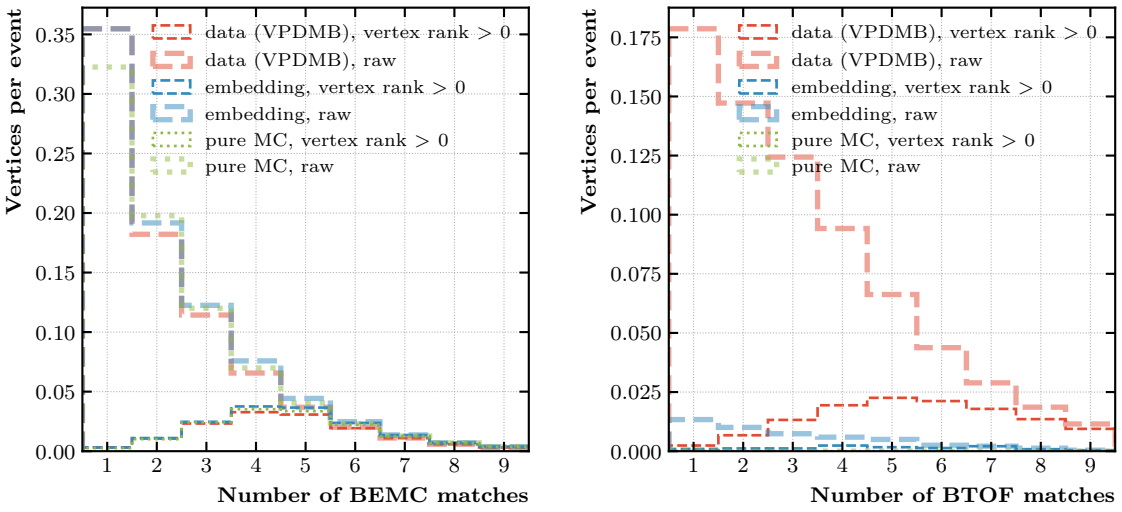
4.2.2 Quality assurance

While the event generation and the detector simulation are made using our understanding of the underlying physics processes, both of these components are imprecisely known, and leave room for discrepancies with data. This situation is complicated by the fact that the comparisons have to be made at the detector level without being able to easily attribute any inconsistency to the right step. For this analysis, the current discrepancies needed to be identified so that they can be corrected or, at least, addressed when estimating the systematic uncertainties for the measurement.

Vertexing

Vertexing efficiency is sensitive to various aspects of the simulation. For example, the vertex ranking depends on matching tracks to the fast detectors, however, unlike for the BEMC, the simulation for the BTOF was not working properly in the Run 12 version of the STAR simulation framework. Results from a vertexing study given in Table 4.3 show good agreement in the vertexing efficiency when a small sample of the Run 12 min-bias data is compared to a STAR simulation. This suggests that the BTOF hits are not significantly affecting the vertex finding.

[1]The reconstructed TPC vertex position would work well here too.



(a) Matches to BEMC

(b) Matches to BTOF

Figure 4.9: Distributions of numbers of TPC tracks matching to BEMC and BTOF reported for vertices reconstructed in the real data and a simulation. This shows potential problems with vertex finding in the Run 12 simulation. The samples are described in the caption of Table 4.3.

# of good vertices in an event:	0	1	2	3	4	5
pp at $\sqrt{s} = 200$ GeV						
data (VPDMB NOBSMD)	85.67%	14.13%	0.20%	0.00%	0.00%	0.00%
embedding ($0 < \hat{p}_T < \infty$)	83.21%	16.59%	0.20%	0.00%	0.00%	0.00%
pure MC ($0 < \hat{p}_T < \infty$)	84.10%	15.90%	0.00%	0.00%	0.00%	0.00%
pp at $\sqrt{s} = 510$ GeV						
data (VPDMB NOBSMD)	53.44%	39.80%	5.97%	0.68%	0.09%	0.02%
embedding ($0 < \hat{p}_T < \infty$)	52.93%	40.56%	5.74%	0.72%	0.05%	0.01%
pure MC ($0 < \hat{p}_T < \infty$)	42.26%	57.48%	0.26%	0.00%	0.00%	0.00%

Table 4.3: Results from a supplementary study of the vertexing efficiency in Run 12 showing what percentage of events have a certain number of "good" (positively ranking) vertices. In the STAR framework, each event is reconstructed with 5 vertices, but only a fraction is considered to be non-pile-up. The embedding sample here is equivalent to the one described in this chapter, except for using an "unbiased" PYTHIA sample (`CKIN(3)`$= 0$). The "pure MC" sample is identical to the "embedding" except for disabled mixing of the zerobias data in the TPC. The remaining details about samples used can be found in [70].

Trigger simulator

Trigger simulation for data was discussed in Section 3.5.1. Figure 4.10 shows distributions similar to Fig. 3.8 but for the simulation. Many of the features of the data plot are reproduced in the plot for the embedding simulation.

Basic jet distributions

A comparison at the level of individual jets allows one to focus the analysis on the quantities that are most relevant to the measurement at hand. The basic jet quantity is its transverse momentum (Fig. 4.11), which gives some rough idea on how the measured cross section will differ from the one predicted by PYTHIA. For the purposes of a binned unfolding procedure it is still important to ensure that the shapes of the jet p_{T} distributions don't differ too drastically, which appears to be the case here. In regards to this measurement, the disrepancy in the cross section normalization is not important, as the overal normalization of the simulation is canceled out in the procedure of unfolding. Another quantity to look at is the direction of the jet thrust axis (Fig. 4.12), which is important to ensure that the simulated jets are sampling the same detector acceptance as the jets in the data.

Another quantity important to the use of the simulation is how well it reproduces the detector resolution of the jets' transverse momentum. There is an astute way to check this, by relying on the fact that in two-jet events momentum conservation should largely force jets to point back to back with an equal in magnitude transverse momenta. Hence, any imbalance in the measured jet p_{T} would have to come from various physical sources of missing p_{T} (e.g., due to soft QCD radiation), but primarily from the measured jet p_{T} at the detector level not exactly matching the jet p_{T} at the particle level, for which the momentum conservation law applies. The di-jet transverse momentum asymmetry was defined as:

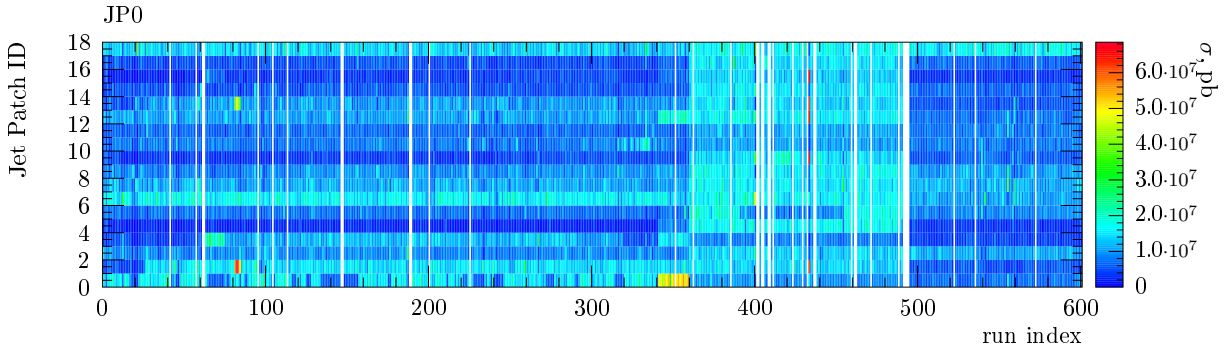
$$A^{(\text{di-jet } p_{\mathrm{T}})} = \sqrt{2}\left|\frac{p_{\mathrm{T};1} - p_{\mathrm{T};2}}{p_{\mathrm{T};1} + p_{\mathrm{T};2}}\right| \tag{4.7}$$

where $p_{\mathrm{T};1}$, $p_{\mathrm{T};2}$ are the transverse momenta of the two jets in the event. The factor of $\sqrt{2}$ is needed to normalize the variance of the $A^{(\text{di-jet } p_{\mathrm{T}})}$, in the limit $\sigma_{p_{\mathrm{T}};1,2} \ll p_{\mathrm{T};1,2}$, to reflect the squared relative p_{T}-resolution $(\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}})^2$ for a single jet. The plot of the average squared deviation from the expected value $\left\langle (A^{(\text{di-jet } p_{\mathrm{T}})} - 0)^2 \right\rangle$ as a function of the characteristic momentum scale of the di-jet is shown in Fig. 4.13. The discrepancy in the variances is very small, on average less than 5%, suggesting very good consistency between the two jet resolutions.

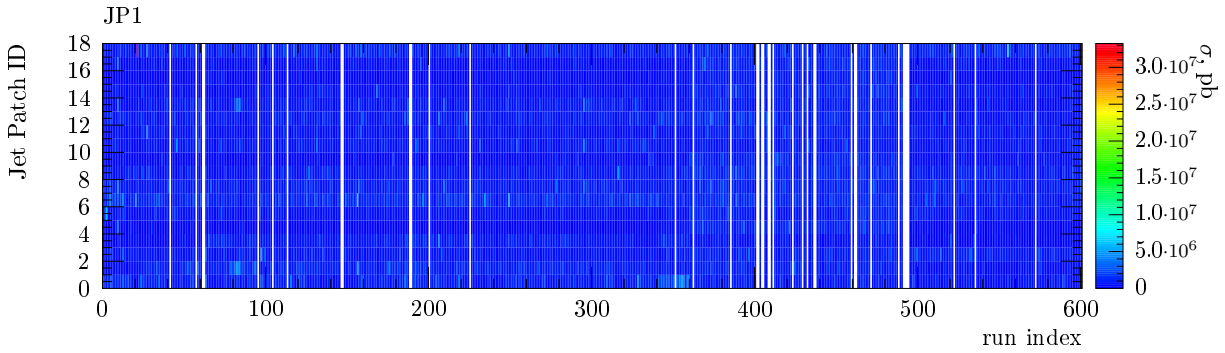
Going deeper into the jet substructure, the detector's response to jets will depend on whether the jets are dominated by tracks or towers. In that regard, we prefer to see good agreement in the distributions of the jet R_{T}. On the plots in Fig. 4.14 it can be seen that there is a preference, at low jet-p_{T} for some extra neutral energy in the simulated sample.

Underlying Event at detector level

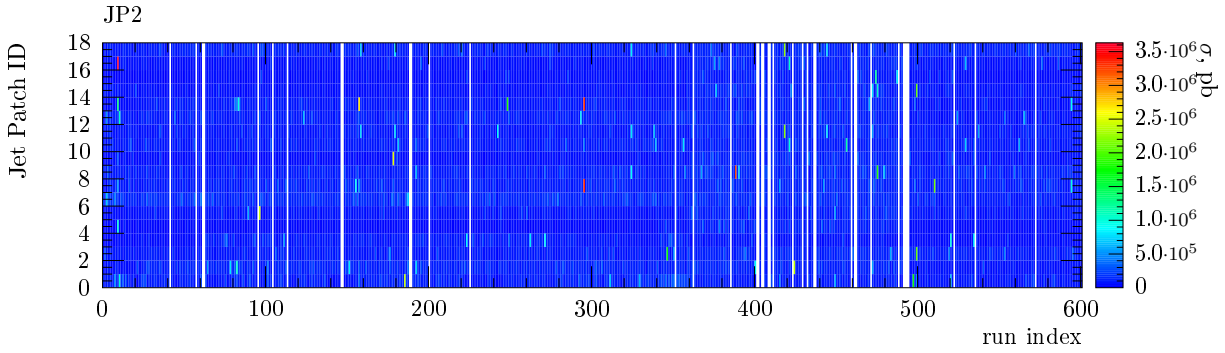
The Underlying Event activity measured in the transverse region cones is an important ingredient for the "UE correction". As was discussed in Section 4.1.1, our PYTHIA tune does



(a) JP0



(b) JP1



(c) JP2

Figure 4.10: Number of events in the embedding simulation for a given run with a given BEMC jet patch with ADC value (simulated offline) above the trigger threshold, shown for the three jet patch triggers used in this analysis.

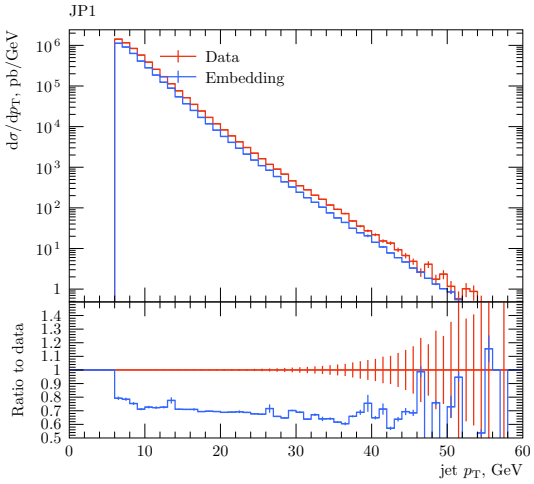


Figure 4.11: Differential detector level cross section for inclusive jets as a function of the jet p_T. Here, the embedding simulation is compared to the actual data.

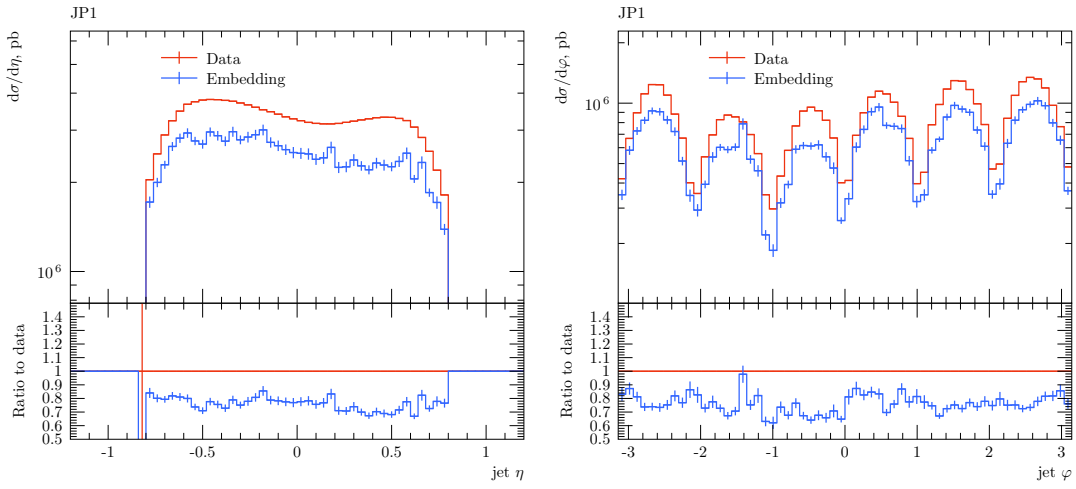


Figure 4.12: Differential detector level cross section for inclusive jets as a function of the jet η and φ.

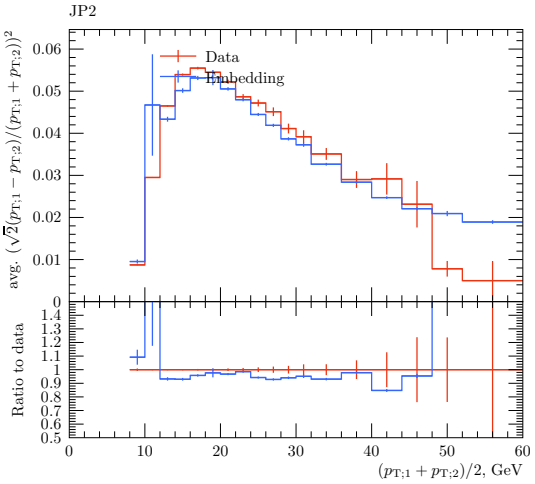


Figure 4.13: Average of the squared asymmetry for jet p_{T} in events with two reconstructed jets as a function of di-jet transverse momentum scale. This variable is constructed as a proxy to the square of the jet momentum resolution of the detector (in this case, the real one and the simulated one are compared).

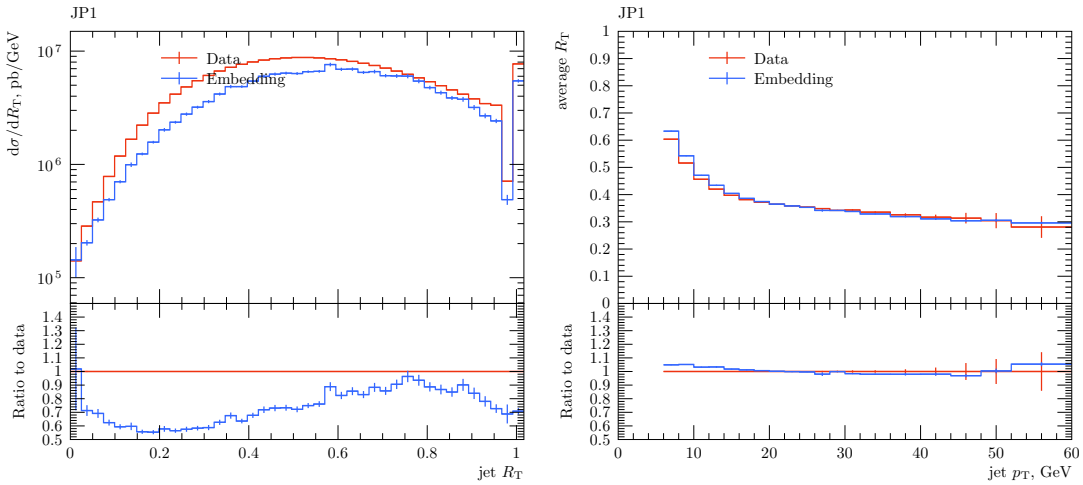


Figure 4.14: Distribution of the jet R_{T} for low-p_{T} jets and the average jet R_{T} plotted as a function of the jet p_{T}.

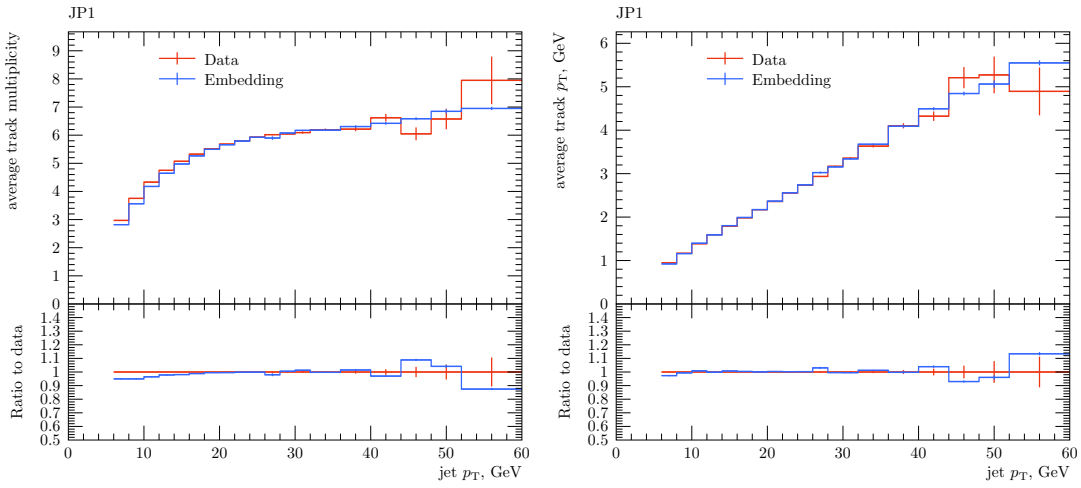


Figure 4.15: Average track multiplicity and track p_T inside each jet.

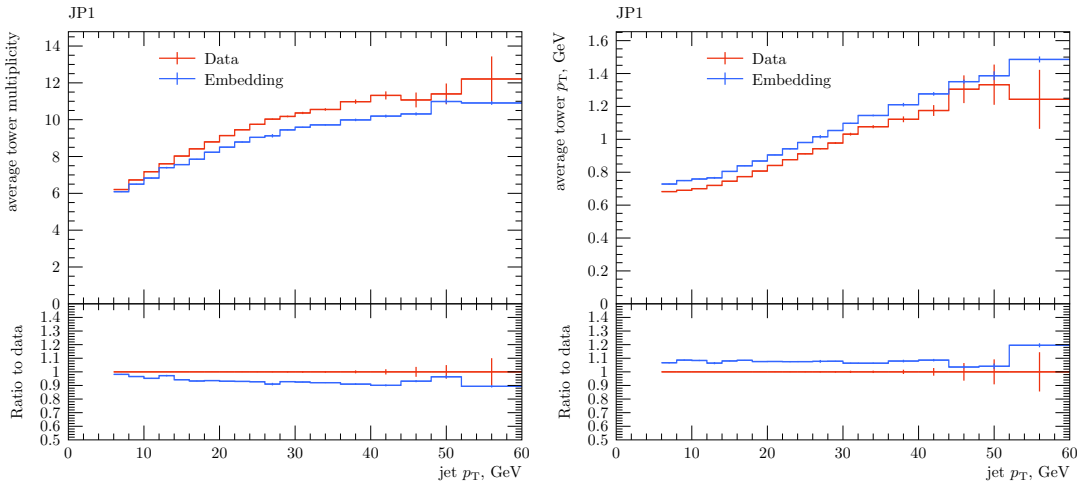


Figure 4.16: Average tower multiplicity and tower p_T inside each jet.

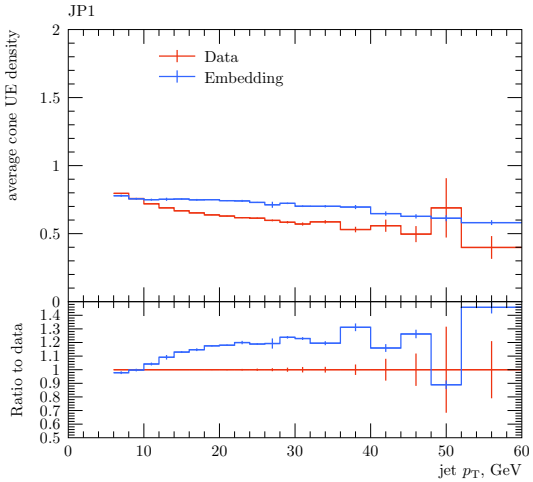


Figure 4.17: Average momentum density in the jet UE cone as a function of the jet p_T.

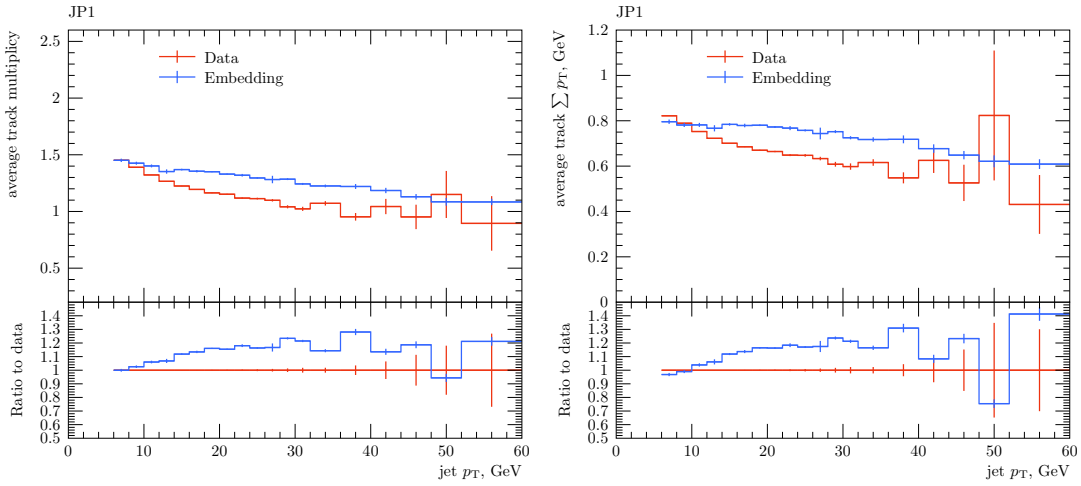


Figure 4.18: Average track momentum density in the jet UE cone as a function of the jet p_{T}.

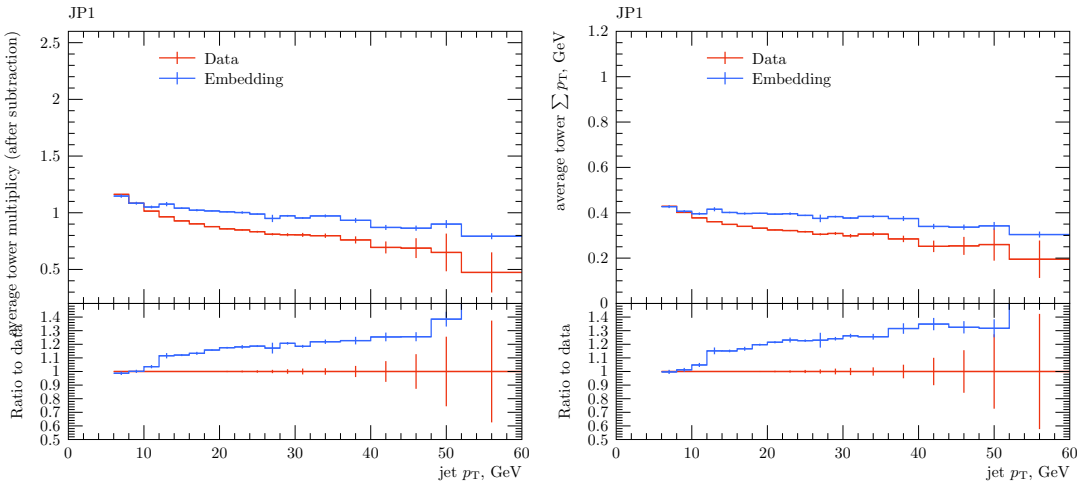
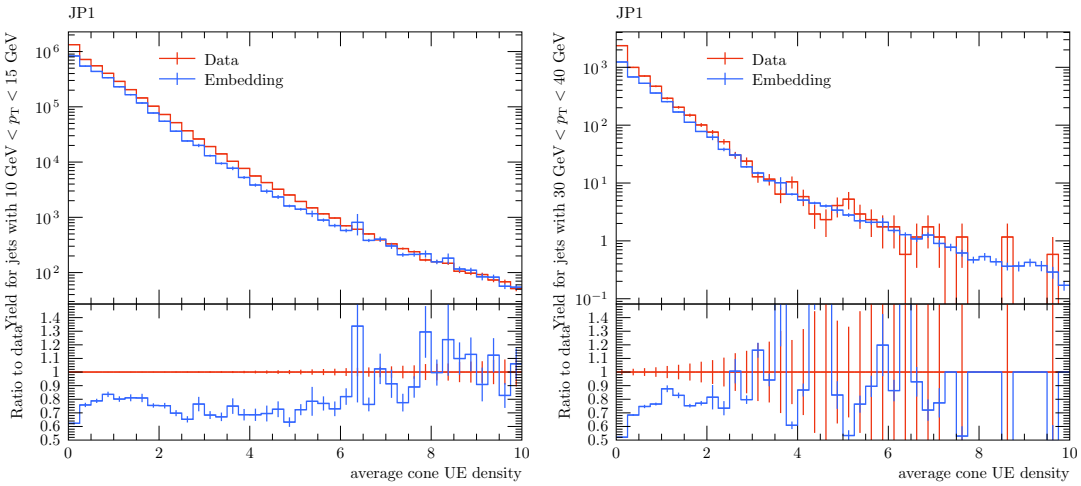


Figure 4.19: Average tower momentum density in the jet UE cone as a function of the jet p_{T}.



(a) $10 <$ jet $p_{\mathrm{T}} < 15$ (b) $30 <$ jet $p_{\mathrm{T}} < 40$

Figure 4.20: Detector level jet cross sections differential in the UE density for two ranges of jet p_{T}.

not reproduce the measured UE, which is consistent with what is seen in Fig. 4.17 at the detector level. Separated contributions of tracks and towers to the cone UE density, shown in Figs. 4.18 and 4.19, prove that the same effect is visible in different detectors, pointing to a physical origin of the effect.

Looking at the first moment of the p_{T} density distribution does not provide a full picture of the Underlying Event activity. The full distributions are displayed in Fig. 4.20. The distributions show an interesting feature, with the data and simulation peaking at ρ_{UE}, but with different heights. This interesting behavior is also seen at the particle level, and at the detector level it is not localized to any special region of detector acceptance. A possible explanation for this behavior is that the peak may be dominated by the events which have no MPI processes occurring.

Chapter 5

Unfolding

Detector jets do not preserve complete and exact information about the original particle jets due to limitations of our detector. Information is lost due to detection inefficiencies and incomplete acceptance, due to limited position, energy and momentum resolution, and also due to an extremely limited detector response to neutrons, neutrinos and some long-lived neutral mesons. The missing information can not be recovered for each individual jet or each individual recorded event. However, we are interested in properties of statistically large ensembles of jets, and, for those, some information can be recovered. In particular, we will be interested in extracting information about the p_T-η distribution of the particle jet yield from the distribution of the detector jet yield, using a process called "unfolding".

Unfolding uses statistical methods grounded in Probability Theory. The basic results for unfolding are unambiguous and can be derived assuming all measured distributions to be known exactly. The notorious challenges in doing the unfolding are attributed to the limited precision in the data (real and simulated) used for measuring the distributions, or to a choice of an unfolding "method". What is often overlooked in the discussion is the importance of having the insight needed to do a proper setup for the process. All aforementioned potential problems will be discussed in this chapter.

5.1 Formulating the goal

We want to first formulate the problem that we want to approach using statistics. The formalism of Probability Theory requires us to define a set of all possible outcomes, then an event is a set of occurred outcomes. We also need to define the probability function P which maps each event onto its probability. Specifically for unfolding, we are dealing with two kinds of outcomes: "cause" outcomes and "effect" outcomes.

To give a concrete example for the jet cross section measurement, let's consider each hard proton-proton interaction to be an event. Though not very realistic, for this example let's also assume that at most one jet can occur in a given collision. The generalization to the case of several jets yields the same basic procedure. In our description, the list of all effect outcomes will look like:

- (effect 1) – a trigger was (mis-)fired and a detector jet with $|\eta| < 0.8$ and 6.9 GeV $\leq$ jet $p_T < 8.2$ GeV can be reconstructed in the event

- (effect 2) – a trigger was (mis-)fired and a detector jet with $|\eta| < 0.8$ and 8.2 GeV $\leq$ jet $p_T < 9.7$ GeV can be reconstructed in the event

  .....

- (effect $n_{\text{eff.}} - 1$) – a trigger was (mis-)fired and a detector jet with $|\eta| < 0.8$ and 37.2 GeV $\leq$ jet $p_T < 44$ GeV can be reconstructed in the event

- (effect $n_{\text{eff.}}$) – a trigger was (mis-)fired and a detector jet with $|\eta| < 0.8$ and 44 GeV $\leq$ jet $p_T < 52$ GeV can be reconstructed in the event

- (effect $n_{\text{eff.}} + 1$) – no trigger was fired, or a trigger was fired, but no detector jet with $|\eta| < 0.8$ and 6.9 GeV $\leq$ jet $p_T < 52$ GeV can be reconstructed in the recorded event

The last effect outcome corresponds to the detector inefficiency towards jet reconstruction as well as the phase space limits used for the applicable data, but also has a sizable contribution of events that legitimately did not have jets of interest. Note that the last "effect" outcome, unlike others, is often not truly observable (one can count the bunch crossings and know the luminosity, but experimentally, it is difficult to precisely count hard proton-proton interactions for normalization).

An example of the set of cause outcomes could look like:

- (cause 1) – jet with $|\eta| < 0.8$ and 6.9 GeV $\leq$ jet $p_T < 8.2$ GeV was produced in the interaction

- (cause 2) – jet with $|\eta| < 0.8$ and 8.2 GeV $\leq$ jet $p_T < 9.7$ GeV was produced in the interaction

  .....

- (cause $n_{\text{cause}} - 1$) – jet with $|\eta| < 0.8$ and 37.2 GeV $\leq$ jet $p_T < 44$ GeV was produced in the interaction

- (cause n_{cause}) – jet with $|\eta| < 0.8$ and 44 GeV $\leq$ jet $p_T < 52$ GeV was produced in the interaction

- (cause $n_{\text{cause}} + 1$) – no jet at all or no jet with $|y| < 0.8$ and 6.9 GeV $\leq$ jet $p_T < 52$ GeV was produced in the hard proton-proton interaction

The last cause outcome contributes to the background "fake jet" production rate, but also to the actual jet-less events. Even with an accurate simulation, we would not expect to be able to measure this one with good precision due to its large overlap with the unmeasurable non-observable outcome mentioned above. The other cause outcomes, which we expect to be measurable, need to be chosen to mostly overlap with the observable outcomes to minimize the model dependence.

The set of all possible events is given by the Cartesian product of the sets defined above:

$$\Omega = \cup_{i=1\ldots(n_{\text{eff.}}+1)} \cup_{j=1\ldots(n_{\text{cause}}+1)} \{(\text{effect } i \wedge \text{cause } j)\} \tag{5.1}$$

Now the problem the unfolding procedure is meant to address can be formulated as follows: given known values of a measured spectrum $b_i = N \cdot P(\text{effect } i)$ for each possible i except the $(n_{\text{eff.}} + 1)$, where N is a normalization (e.g. number of trials) that is generally unknown, to estimate a truth spectrum $x_j = N \cdot P(\text{cause } j)$ for each possible j except the $(n_{\text{cause}} + 1)$.

5.2 Response matrix

Working within definitions from the previous section, one can immediately write down the following two relations between these probabilities:

$$\begin{aligned} P(\text{effect } i) &\equiv \sum_{j=1\ldots(n_{\text{cause}}+1)} P(\text{cause } j \wedge \text{effect } i) = \\ &= \sum_{j=1\ldots(n_{\text{cause}}+1)} P(\text{effect } i \,|\, \text{cause } j) \cdot P(\text{cause } j) \end{aligned} \tag{5.2}$$

$$\begin{aligned} P(\text{cause } j) &\equiv \sum_{i=1\ldots(n_{\text{eff.}}+1)} P(\text{cause } j \wedge \text{effect } i) = \\ &= \sum_{i=1\ldots(n_{\text{eff.}}+1)} P(\text{cause } j \,|\, \text{effect } i) \cdot P(\text{effect } i) \end{aligned} \tag{5.3}$$

Both of these relations connect the values of $P(\text{cause } j)$ and $P(\text{effect } i)$ just like we need. A full Monte Carlo simulation using the Pythia event generator and the detector simulation provides us with an estimate for the value of all probabilities $\{P(\text{cause } j \wedge \text{effect } i)\}$, albeit, those probabilities are given for a world where the $\{P(\text{cause } j)\}$ is already fixed by the event generator. That information is factored out in the set of values for $P(\text{effect } i \,|\, \text{cause } j)$, but not for $P(\text{cause } j \,|\, \text{effect } i)$. One is not able to fix both values at the same time, or equivalently use both equations. Hence, Eq. (5.2) is immediately preferred over Eq. (5.3). In fact, Eq. (5.3) may intuitively appear as correct, but leads to invalid measurements.

Rewriting Eq. (5.2) in terms of x_j and b_i:

$$b_i = \sum_{j \in \{1\ldots n_{\text{cause}}\}} \hat{A}_{ij} x_j + b_{0;i} \tag{5.4}$$

where $\hat{A}_{ij} \equiv P(\text{effect } i \,|\, \text{cause } j)$ is the response matrix that can be estimated directly from the simulation and $b_{0;i} \equiv N \cdot P(\text{effect } i \,|\, \text{cause } (n_{\text{cause}} + 1)) \cdot P(\text{cause } (n_{\text{cause}} + 1))$ is the background, which is typically hard to model and measure. In our case, it is modelled as multiplicative:

$$b_{0;i} = \frac{b_{0;i}^{\text{simu.}}}{b_i^{\text{simu.}}} b_i \tag{5.5}$$

where the ratio $b_{0;i}^{\text{simu.}} / b_i^{\text{simu.}}$ can be called a "background rate".

Then one can define $M_{ij} \equiv \left(1 - \frac{b_{0;i}^{\text{simu.}}}{b_i^{\text{simu.}}}\right)^{-1} \hat{A}_{ij}$ to include both the background and the efficiency effects in one matrix.

5.3 Linear system solution

Let's look again at the system from Eq. (5.4), now in a matrix form:

$$M\,x = b \tag{5.6}$$

The dimension of x is n_{cause} and the dimension of b is $n_{\text{eff.}}$. In cases when $n_{\text{cause}} = n_{\text{eff.}}$, the matrix M is square, one can use a simple matrix inverse[1] to solve Eq. (5.6):

$$x = M^{-1}b \tag{5.7}$$

In cases when $n_{\text{cause}} > n_{\text{eff.}}$ the linear system will be formally overdetermined, which poses little problem as the system can be solved in the least squares sense, i.e. as a solution to:

$$\chi^2 = (M\,x - b)^T(M\,x - b) = \min \tag{5.8}$$

In the case where $n_{\text{eff.}} \leq n_{\text{cause}}$, these two formulations are equivalent. (And the case when $n_{\text{eff.}} < n_{\text{cause}}$ is undesirable as a solution to such a problem will not be uniquely defined)

The solution to the minimization problem is found as

$$0 = \frac{\partial \chi^2}{\partial x^T} = M^T(Mx - b) = M^T M\,x - M^T b \tag{5.9}$$

a solution to which is

$$x = (M^T M)^{-1} M^T b \tag{5.10}$$

The solution of this form is connected to the "normal equation".

In the case of the jet p_{T} spectrum, it is originally continuous, but needs to be binned for unfolding. Then the case of $n_{\text{cause}} > n_{\text{eff.}}$ is not beneficial as it increases the number of elements in M, which introduces additional statistical fluctuations and puts the accuracy of the simulation to a more stringent test.

5.4 Regularization

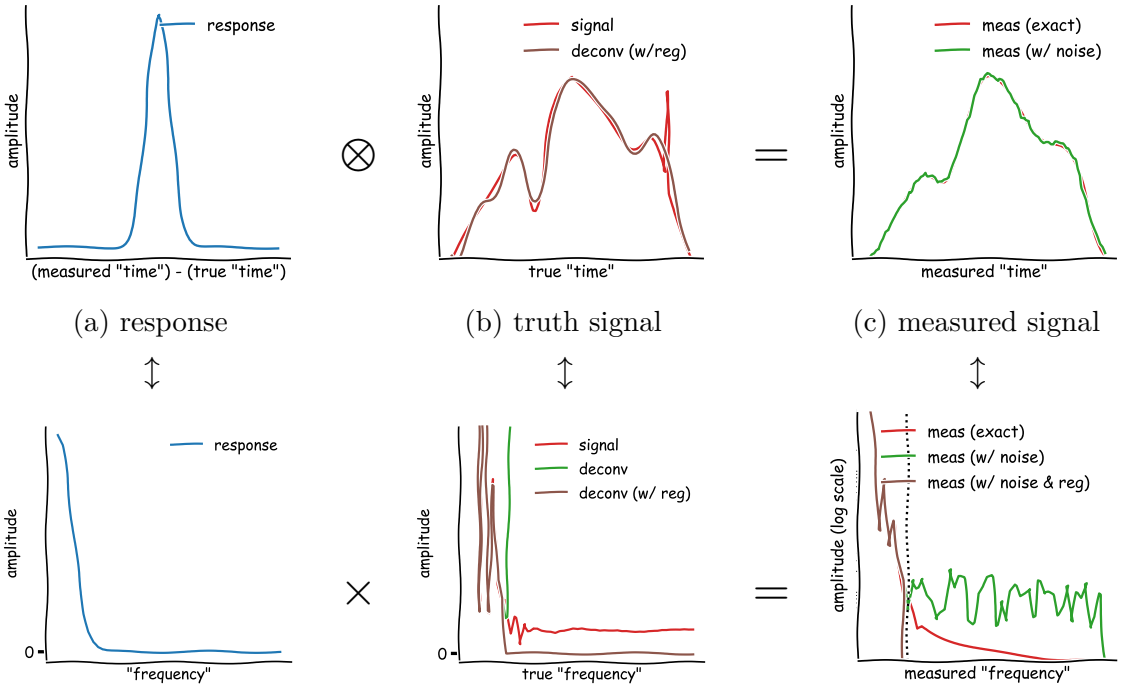
Certain cases may or may not warrant introduction of regularization into the unfolding. The cases are best illustrated with toy examples to help with building up some intuition.

5.4.1 Example: signal processing

A simple toy example for unfolding can be constructed by imagining a measurement of a signal $f(t)$ (Fig. 5.1b) that is subject to a linear response that is t-invariant. That process can be described in terms of convolution with a response $R(t)$:

$$(f_{\text{truth}} \otimes R)(t) \equiv \int f_{\text{truth}}(\tau)\, R(t-\tau)\, d\tau = f_{\text{measured}}(t) \tag{5.11}$$

[1] It is not rare that potential non-invertibility of M is mentioned. That should be taken as a signal that something is wrong with the setup – the detector is not capable of doing the measurement. A typical case for this may be a row that is all zero because of incorrectly setup bin boundaries.



(a) response (b) truth signal (c) measured signal

(d) response (freq. domain) (e) truth signal (freq. domain) (f) measured signal (freq. domain)

Figure 5.1: A cartoon illustrating unfolding of a continuous signal smeared with a Gaussian response. The measurement acquires some noise background, and an attempt to unfold it fails without a frequency cut off.

To do the measurement we would need to solve this equation for $f_{\text{truth}}(t)$. An easy way to do that is to consider the same equation in Fourier space where convolution "$\otimes$" becomes a simple number multiplication " $\cdot$ ":

$$\tilde{f}_{\text{truth}}(\omega) \cdot \tilde{R}(\omega) = \tilde{f}_{\text{measured}}(\omega) \tag{5.12}$$

hence the solution is simply

$$\tilde{f}_{\text{truth}}(\omega) = \frac{\tilde{f}_{\text{measured}}(\omega)}{\tilde{R}(\omega)} \tag{5.13}$$

Typically, the response $R(t)$ will have some kind of "bell" shape; to be specific we assume it to be a Gaussian (Fig. 5.1a):

$$R(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{5.14}$$

consequently the R-related term in the solution in Eq. (5.13) is

$$\frac{1}{\tilde{R}(\omega)} = \frac{\sqrt{2\pi}}{\sigma} \exp\left(\frac{\sigma^2\omega^2}{2}\right) \tag{5.15}$$

We can immediately see the problem here, which is that for large values of ω this multiplier grows to be very large. This, in turn, leads to a strong amplification of the high-frequency part of the measured signal (Fig. 5.1f). In the ideal situation where the measurement is truly governed by Eq. (5.11), there would be no issue, as the exact spectrum will decrease with frequency in accordance with the response. The real world measurements, however, will often come with high-frequency noise. In signal processing it would typically come from stochastic processes, and for our counting experiment it will come from statistical fluctuations between the histogram bins.

A reasonable procedure for regularization here could involve identifying a noise floor for the signal, to determine a cutoff frequency beyond which to ignore the measured values.

Another issue that one can think of in relation to this toy example is what happens when the response assumed during the unfolding differs from the actual response. It's easy to see that extra high frequency contributions arise if the width of the "smearing" parameter σ is overestimated. Referring back to Fig. 4.13, it's probably for the best that the jet p_{T} resolution in the simulation appears to be slightly better than in the data.

5.4.2 Example: discrete spectrum unfolding

In the previous section an example with a continuous spectrum was considered. In the case of our jet counting experiment, the events with somewhat different jet p_T values need to be grouped into bins for the measured total counts to have good statistical precision. The discrete case can be described by the linear system Eq. (5.6) that is of the same nature as Eq. (5.11).

We will now consider a toy example from [71], by imagining a detector described by a response matrix

$$M = \frac{1}{2}\begin{pmatrix} 1+\epsilon & 1-\epsilon \\ 1-\epsilon & 1+\epsilon \end{pmatrix}, \qquad \text{where } \epsilon \ll 1 \tag{5.16}$$

The solution for Eq. (5.6) is

$$x \equiv \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{b_1 + b_2}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{b_1 - b_2}{2\epsilon} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{5.17}$$

Let's analyze it. Looking at the first term in Eq. (5.17), we can conclude that a detector described by response from Eq. (5.16) has no problem in measuring an average between x_1 and x_2, as it is just $(x_1 + x_2)/2 = (b_1 + b_2)/2$. However, if we are interested in a difference between x_1 and x_2, the solution for that is $x_1 - x_2 = (b_1 - b_2)/\epsilon$, and it is proportional to a large number ϵ^{-1}. Unless the uncertainty for the measurement of $(b_1 - b_2)$ is sufficiently low (specifically $\delta\{b_1 - b_2\}/\epsilon \ll (b_1 + b_2)$), the direct inversion solution will yield an error-dominated result for the individual values of x_1 and x_2, although that error is a correlated one.

While the measured values are technically valid within the estimated uncertainties, they may be deemed as unpresentable, or even appear useless. There is no solution to this problem that can be made without making some additional assumptions. In the absence of these, one should just admit that the detector described by Eq. (5.16) is lacking the resolution needed to measure the value of $x_1 - x_2$.

In practice the dimensions of vectors b and x are often higher than 2, and it would be nice to be able to perform a similar analysis in the case of arbitrary dimensions. In order to do that, one can consider a Singular Value Decomposition of the matrix M:

$$M = U\,S\,V^T \tag{5.18}$$

where U and V are orthogonal matrices and S is a diagonal[2] matrix. The values on the diagonal of S are non-negative and ordered in descending order.

For our example in Eq. (5.16), we get:

$$M = \frac{1}{2} \begin{pmatrix} 1+\epsilon & 1-\epsilon \\ 1-\epsilon & 1+\epsilon \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^T \tag{5.19}$$

The columns of the orthogonal matrices U and V, by definition, form two (generally, different) orthonormal bases: left and right. Those have a one-to-one correspondence between each other and the diagonal elements of S. In a typical measurement, the matrix M is of a kind that smears features of the distributions represented by vectors x and b, so a pair of respective left and right basis vectors with sharper features (most often coming as more oscillation) will generally have lesser singular values than a pair with less oscillation (see example in Fig. 5.2).

When taking the inverse of M, $M^{-1} = V S^{-1} U^T$ leads to a simple exchange of left and right bases and, more importantly, the smallest singular values of M now become the largest singular values of M^{-1}. They are responsible for amplification of the statistical fluctuations that come as high-frequency "noise"[3] to the measured spectrum b. This is similar to what we saw when discussing Eq. (5.15).

[2] If M is not square, S is also not square, but still diagonal, and U and V are square.

[3] What is noise here will be determined by the orthogonal basis made up of columns of V. Generally for a smearing matrix the smaller singular values will correspond to most "fluctuating" vectors.

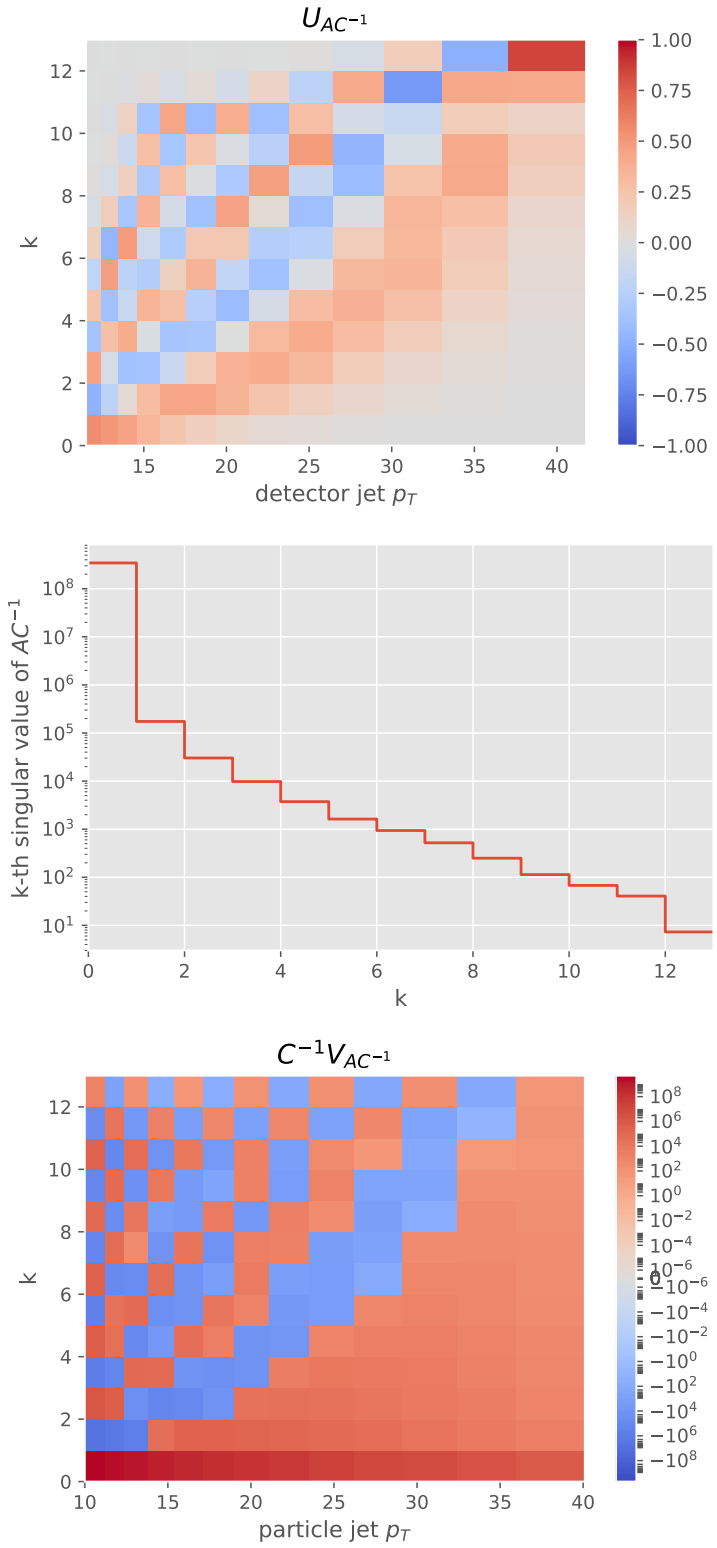


Figure 5.2: An example of SVD decomposition for a smearing matrix. The higher level oscillations correspond to a higher wave vector k with a lower singular value s_k.

5.4.3 Tikhonov regularization

Instead of solving Eq. (5.6) via matrix inversion or by minimization $x = \arg\min_x (M\,x - b)^2$, one could assume a prior for x. In that case the sought solution may acquire a bias term and become

$$x = \arg\min_x ((M\,x - b)^2 + \tau (D\,x)^2) \tag{5.20}$$

Specifically, when looking to dampen the oscillating terms, one would be looking to choose D representing a second derivative in the bin index (in this case, applied to the ratio of truth spectra over the prior spectra):

$$D = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & ... \\ -1 & 2 & -1 & 0 & ... & 0 \\ 0 & -1 & 2 & ... & 0 & 0 \\ 0 & 0 & ... & 2 & -1 & 0 \\ 0 & ... & 0 & -1 & 2 & -1 \\ ... & 0 & 0 & 0 & -1 & 1 \end{pmatrix} \mathrm{diag}\left\{\frac{1}{x_{\mathrm{prior}}}\right\} \tag{5.21}$$

The solution to Eq. (5.20), if it exists, is found amongst the stationary points of the functional. Those are given by the "normal equation" solution:

$$x = (M^T M + \tau D^T D)^{-1} M^T b \tag{5.22}$$

Tikhonov regularization equivalence to a prior

Some motivation for the form of Tikhonov regularization can be found in connection to Bayesian statistics approach [72].

The matter of choice of a prior in the Bayesian approach is subjective. For example, we may have a prior assumption that the unfolded spectrum would be sufficiently smooth. The corresponding prior probability could be written as something like

$$P(\vec{x}) = \frac{\det C}{\sqrt{(2\pi)^3}} \exp\left(-\frac{(\vec{x} - \vec{x}_{\mathrm{ref}})^{\mathrm{T}} C^{\mathrm{T}} C (\vec{x} - \vec{x}_{\mathrm{ref}})}{2}\right) \tag{5.23}$$

The likelihood of observing the measured spectrum $\vec{b}$ given the truth spectrum $\vec{x}$ is then given by

$$P(\vec{b}|\vec{x}) = \frac{1}{\sqrt{(2\pi)^3 \det B}} \exp\left(-\frac{(b - A\vec{x})^T B^{-1} (b - A\vec{x})}{2}\right) \tag{5.24}$$

The Bayesian inference then suggests a following posterior log probability for the truth spectrum to be $\vec{x}$ given measured spectrum $\vec{b}$:

$$\begin{aligned} \log(P(\vec{x})|P(\vec{b})) &= \log(P(\vec{b}|\vec{x})) + log(P(\vec{x})) - log(P(\vec{b})) \\ &= (b - A\vec{x})^T B^{-1} (b - A\vec{x}) + (\vec{x} - \vec{x}_{\mathrm{ref}})^{\mathrm{T}} C^{\mathrm{T}} C (\vec{x} - \vec{x}_{\mathrm{ref}}) + \mathrm{const}(\vec{x}) \end{aligned} \tag{5.25}$$

This demonstrates how Tikhonov regularization is equivalent to having a certain Bayesian prior.

Tuning Tikhonov regularization based on SVD analysis

The question remains about the value of the regularization strength parameter τ in Eq. (5.20). To help with the choice, it is useful to first understand its role.

For this, Eq. (5.20) can be solved in a way alternative to Eq. (5.22), by first making a substitution $y \to Dx$:

$$x = D^{-1} \arg\min_y((MD^{-1}y - b)^2 + \tau y^2) \tag{5.26}$$

And then taking the Singular Value Decomposition (SVD) of the matrix MD^{-1}, $USV^T = MD^{-1}$ where U and V are orthogonal matrices and S is the diagonal matrix $S = diag(\{s_k\})$.

$$\begin{aligned} x &= D^{-1} \arg\min_y((USV^T y - b)^2 + \tau y^2) && (5.27)\\ &= D^{-1} \arg\min_y((USV^T y - b)^2 + \tau y^T V V^T y) && (5.28)\\ &= D^{-1} V \arg\min_z((USz - b)^2 + \tau z^2) && (5.29)\\ &= D^{-1} V \arg\min_z(z^T(S^2 + \tau)z + b^T b - 2z^T S U^T b) && (5.30)\\ &= D^{-1} V \frac{S}{S^2 + \tau} U^T b && (5.31) \end{aligned}$$

Here another substitution $z = V^T y$ is used before picking a stationary point as the minimum. The resulting alternative expression for the solution from Eq. (5.22) allows one to explicitly see the effect that the Tikhonov regularization has on the result of unfolding. Namely, vectors corresponding to singular values $s_k^2 \ll \tau$ are now not amplified with the full force $1/s_k$, but instead dampened by a factor $\approx s_k^2/\tau$. On the other hand, the contributions corresponding to the singular values such that $s_k^2 \gg \tau$ remain mostly unchanged compared to the unfolding without regularization.

Further analysis can be performed assuming infinite precision in the estimation of M, but finite precision of the data b. The statistical uncertainty of b can be represented by its covariance matrix:

$$B \equiv \mathrm{cov}(b, b^T) \tag{5.32}$$

In light of Eq. (5.27), what presently is of more interest is not the uncertainty on b, but the uncertainty on the vector $d = U^T b$, with its corresponding covariance matrix:

$$D \equiv \mathrm{cov}(d, d^T) = U^T B U \tag{5.33}$$

In the method presented in [71], at this point B would be represented by an identity matrix, hence D would be as well. This is not essential to the method though. The d_i can be compared to the diagonal elements $\sqrt{D_{ii}}$ to determine an "effective rank". If the index i (equivalent of the wave number) is ordered by decreasing singular values, one could expect that for low singular values (at high "frequency"[4]) the statistical uncertainty $\delta\{d_i\} = \sqrt{D_{ii}}$ will be larger than d_i itself. Then one can define a boundary, a specific index that would

[4]It may be important that the SVD decomposition is not performed on the smearing matrix M. The MD^{-1} may act not as a smearing matrix in certain cases, so this form of Tikhonov regularization may not give reasonable results.

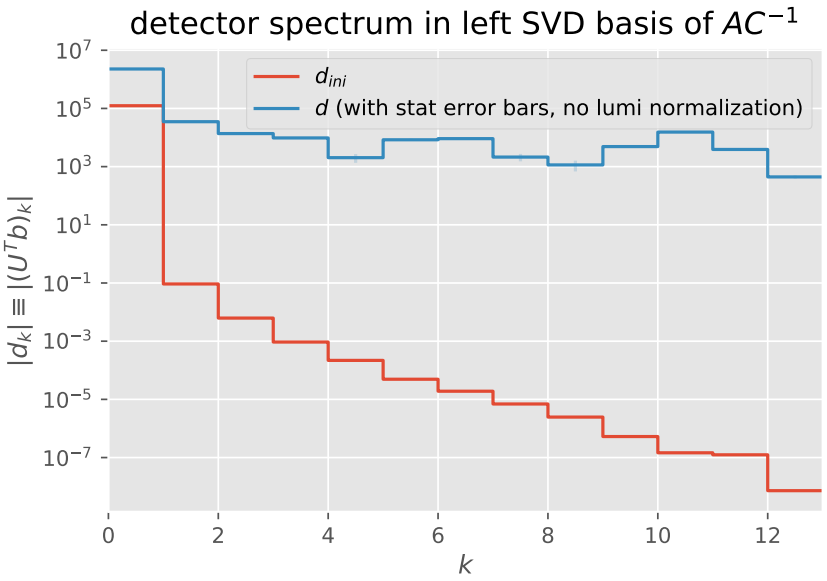


Figure 5.3: Plot of a p_{T} jet spectrum in the SVD basis of the data d_i and the simulation d_i^{ini}. The error bars are too small to be visible. The d_i^{ini} should be zero for all $i \neq 0$, but this is due to a numerical artifact.

separate the values that mostly satify $\delta\{d_i\} < d_i$ and those that mostly satisfy $\delta\{d_i\} > d_i$, which can be called the effective rank. Then knowing that the effective rank is n, the τ should be set to s_n^2. In practice, this step is often ignored completely. For example, the RooUnfold package doesn't facilitate inspection of d_i, so users set an effective rank to a value that makes their results look fine.

In practical application, the present jet cross section measurement had plenty of data statistics available so that $\delta\{d_i\} \ll d_i$ (Fig. 5.3), so there was no reason to apply the Tikhonov regularization to it.

Use of inverse covariance matrix term

In [71], the system solved is different from Eq. (5.20) due to the presence of the inverse of the covariance matrix B:

$$x = \arg\min_x((M\,x - b)^T B^{-1}(M\,x - b) + \tau(D\,x)^2) \tag{5.34}$$

The solution for this can be achieved using, for example, the Cholesky decomposition $B^{-1} = L^T L$ and substituting $M \to LM$ and $b \to Lb$ in the solutions found above. This use of Eq. (5.34) would typically be a reasonable thing to do for an overdetermined system of equations (i.e., when M is not a square matrix and $\dim(b) > \dim(x)$) in the least-squares sense. However, in the case of the exact solutions for a square system there is no such advantage.

There is, however, an interference with the regularization term in either of these cases.

To see that, let's consider the following example:

$$A = \frac{1}{2}\begin{pmatrix} 1+\epsilon & 1-\epsilon & 0 & 0 \\ 1-\epsilon & 1+\epsilon & 0 & 0 \\ 0 & 0 & 1+\mu & 1-\mu \\ 0 & 0 & 1-\mu & 1+\mu \end{pmatrix}, \text{ with } \epsilon, \mu \ll 1 \tag{5.35}$$

$$b \simeq \begin{pmatrix} 100 \\ 100 \\ 1 \\ 1 \end{pmatrix}, \qquad B = \text{diag}(b) \tag{5.36}$$

One can check that the effect of regularization in this case is that the difference $b_1 - b_2$ is suppressed by a factor $\frac{\epsilon^2}{\epsilon^2+100\tau}$ whereas $b_3 - b_4$ only by a factor $\frac{\mu^2}{\mu^2+\tau}$. The issue here is that the vector that can achieve better statistical precision got disproportionately penalized by the regularization.

5.4.4 "Iterative Bayesian unfolding"

A procedure described in [73] first introduces an iteration step of the Richardson-Lucy algorithm [74, 75]:

$$P(\text{cause } i) = \frac{1}{\sum_{k=1\ldots n_{\text{eff.}}} P(\text{effect } k \,|\, \text{cause } i)} \sum_{j=1\ldots n_{\text{eff.}}} P(\text{cause } i \,|\, \text{effect } j)\, P(\text{effect } j) \tag{5.37}$$

where

$$P(\text{cause } i \,|\, \text{effect } j) = \frac{P(\text{effect } j \,|\, \text{cause } i)\, P_0(\text{cause } i)}{\sum_{k=1\ldots n_{\text{cause}}+1} P(\text{effect } j \,|\, \text{cause } k)\, P_0(\text{cause } k)} \tag{5.38}$$

The paper [73] makes claims about the advantages of this procedure, but leaves out the discussion of its connection to the challenges of unfolding and how it is supposed to address those. Seeing connections to other solutions would also help in understanding this procedure.

In regard to the claim of having a mathematical foundation, it is worth noting that it's not obvious at all that this is directly related to the Bayes's theorem that is so easily proven in Probability Theory. The reason for this is that the prior probability $P_0(\text{cause } k)$ is not the same as the probability after the iteration step $P(\text{cause } i)$. The Bayesian inference that relies on Bayes's theorem looks a bit different, and the same author had later published a different paper which proposes another procedure that actually uses it [76].

It also doesn't clearly spell out how a user should choose the number of iterations to apply. If that number is a regularization parameter, the answer could be one, but the paper mentions applying a custom smearing between the steps and hints at looking at the converging results, so it could be that the number of steps is not meant for regularization after all.

Convergence point

Equation (5.2) looks vastly different from Eq. (5.37), so a natural question arises: Does this mean that the answer given by this procedure would be different from the answer given

by solving Eq. (5.6)? If the solutions were different, that could imply that there are some differences in the assumptions taken. This motivates an investigation.

Consider the iteration step Eq. (5.37) as a mapping $f : R_+^N \mapsto R_+^N$ such that:[5]

$$f_i(\vec{x}) = \frac{1}{\epsilon_i} \sum_j \frac{M_{ji} x_i}{\sum_k M_{jk} x_k} b_j, \qquad \text{where } \epsilon_i = \sum_k M_{ki} \tag{5.39}$$

This expression is simply a compact representation of the iterative procedure written in the same notation as was used for dealing with response matrices.

$$\begin{aligned} f_i(\vec{x}^0 = M^{-1}\vec{b}) &= \frac{1}{\epsilon_i} \sum_j \frac{M_{ji} x_i^0}{\sum_k M_{jk} \sum_l M_{kl}^{-1} b_l} b_j \\ &= \frac{1}{\epsilon_i} \sum_j \frac{M_{ji} x_i^0}{b_j} b_j \\ &= \frac{1}{\epsilon_i} \epsilon_i x_i^0 \\ &= x_i^0 \end{aligned} \tag{5.40}$$

This proves that $\vec{x}^0$ is a fixed-point of f, or in other words, if the prior spectrum was equal to a result of unfolding using Eq. (5.7), then the iteration of this procedure will not change it. This suggests that consecutive applications of the iteration must be converging in the limit of infinite applications to the $\vec{x}^0$ for any other possible starting prior. To prove that mathematically, we are missing just one component needed to apply the Banach fixed-point theorem [77]: that f must be a "contraction mapping" defined on a complete metric space. For a mapping to be a contraction mapping it has to satisfy the criterion:

$$d(f(x), f(y)) \leq q\, d(x, y), \text{ for all } x \text{ and } y \tag{5.41}$$

where d is a metric (e.g., Euclidean distance) and q is a constant such that $q \in [0, 1)$.

Before trying to analytically prove that f is a contraction map, it's worth investigating it in a numerical experiment. Figure 5.4 shows an attempt at it. It would appear that at least in some domain of x the mapping is non-contracting, which in itself doesn't preclude the convergence.

Convergence speed

A common approach in applying Iterative Bayesian unfolding is to set up a "closure test" by slightly modifying a known truth distribution, which can be then mapped onto a measured distribution using the response matrix. Then three or four iterations of the method are applied to some preselected prior and the "toy" measurement, to see if it converges to the original truth.

[5] Here, to use M instead of $\hat{A}$, so that the reader is not confused, it is also assumed that the background cause ($n_{\text{cause}} + 1$) can not be measured and is omitted in the $\sum_k$ summation.

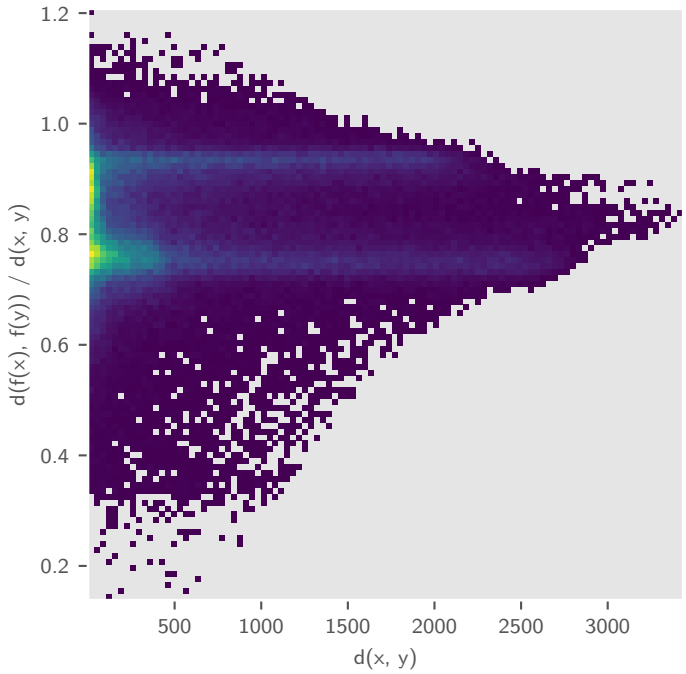


Figure 5.4: Monte Carlo estimate of the contraction mapping criteria for the Iterative Bayesian step using the Euclidean metric. A real-world 12-dimensional response matrix was used with the prior x sampled around the true prior, and y sampled close to x (to keep the distance small).

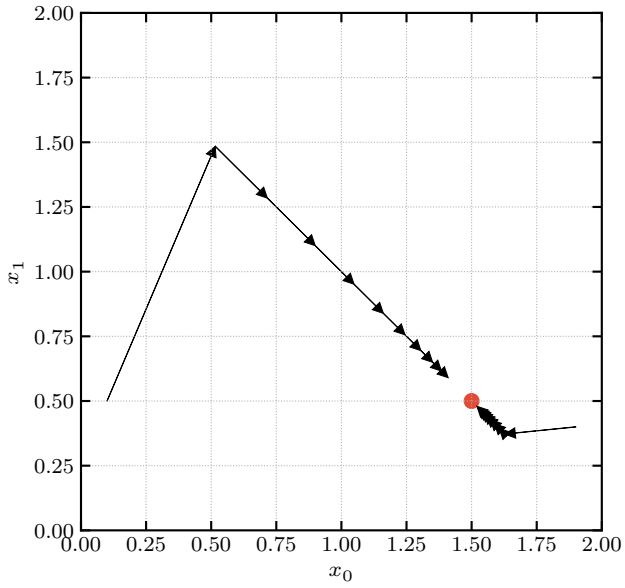


Figure 5.5: Two example step-by-step trajectories of Iterative Bayesian step convergence for the response given in Eq. (5.16) with $\epsilon = 0.5$. The big red dot marks the location of the exact solution.

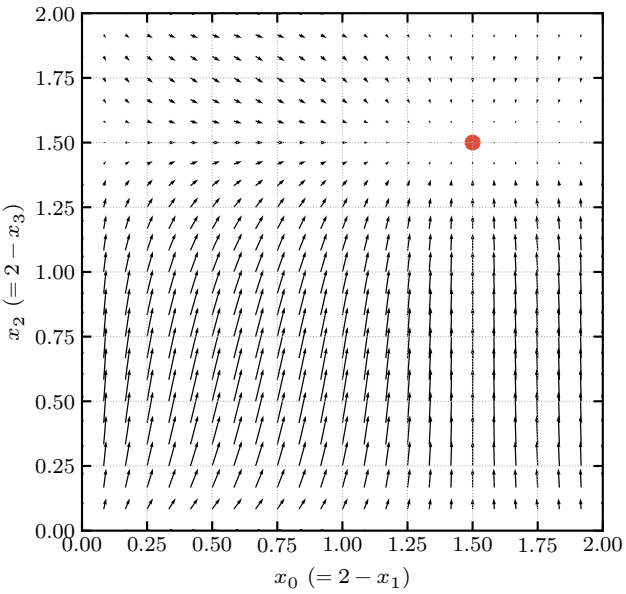


Figure 5.6: Iterative Bayesian step presented as a map. The response from Eq. (5.35) is used with $\epsilon = 0.2$ and $\mu = 0.4$. Only two variables are displayed, as the other two are fixed by the normalization. The big red dot marks the location of the exact solution.

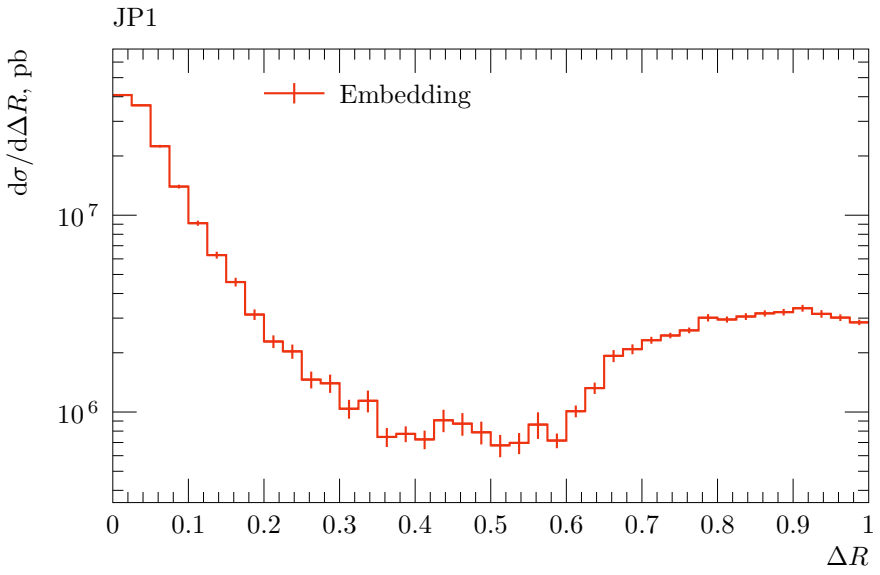


Figure 5.7: Distribution of $\Delta R_{i,j}$ (defined in Eq. (5.42)) for all possible particle and detector jet pairs.

Despite its apparent lack of rigor, this procedure should be useful enough for most cases. However, in a real world measurement the unfolding is almost never done only once. One often has a need to do several auxiliary unfoldings to propagate the detector systematic uncertainties, which may or may not create some unexpected results.

To help understand the convergence properties, a few tests were made using the toy example from Eq. (5.16). Looking at Fig. 5.5, one can see that the first iteration always locks the normalization to match the total number of events with the efficiency correction. The resulting convergence performance is not very impressive even for a large value of $\epsilon = 0.5$.

Another example with two 2×2 matrices put to the test is shown in Fig. 5.6. This clearly shows that convergence for the element with the larger gap is faster.

5.5 Unfolding the inclusive jet cross section

5.5.1 Matching algorithm

When dealing with events that potentially can have multiple jets at the particle level and at the detector level, a question arises on how to establish the causality between the two levels of jets. We make an assumption that one particle jet causes one and only one detector jet. A measure needs to be introduced to quantify how likely it is for one jet to cause another. Normally that could come from the simulation keeping track of which particles caused which detector response. However, in STAR simulation this is not implemented for towers. Hence, an alternative measure of distance between jets in η-φ space was chosen:

$$\Delta R_{i,j} \equiv \sqrt{(\text{detector } \eta_i - \text{particle } \eta_j)^2 + (\text{detector } \varphi_i - \text{particle } \varphi_j)^2} \tag{5.42}$$

The algorithm to match up the jets consists of the following steps that are applied to each event:

1. Select particle level jets that satisfy particle level cuts and detector level jets that satisfy detector level cuts.

2. Take the Cartesian product of these two sets. For each pair of particle and detector jets from that product, assign a distance ΔR according to Eq. (5.42).

3. Select the pair with minimum value of ΔR. If $\Delta R < 0.2$ (cut determined from the distribution in Fig. 5.7), call them "matching". If not, skip to the next event.

4. Remove all pairs that have the same particle or detector jet as a pair from the previous step.

5. Go back to step 3 until the list of pairs is empty.

The result of this algorithm was a list of detector-particle jet pairs that were matched, and lists of the detector jets and of the particle jets that were not matched.

5.5.2 Matching rates

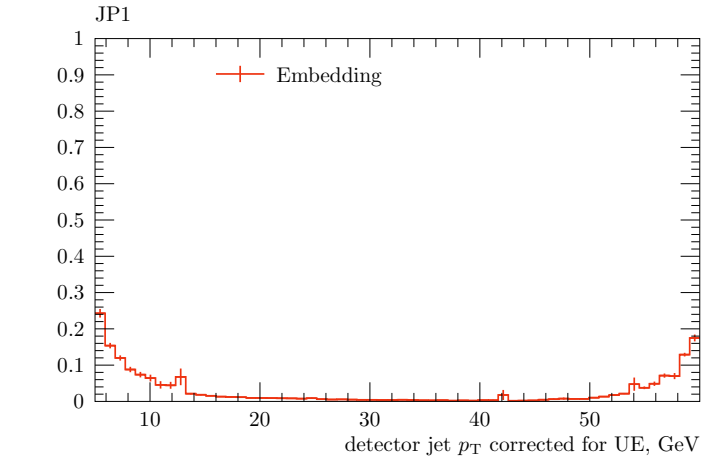
The two matching rates can be estimated from the procedure:

- The fraction of detector jets that were matched to a particle jet
- The fraction of particle jets that were matched to a detector jet

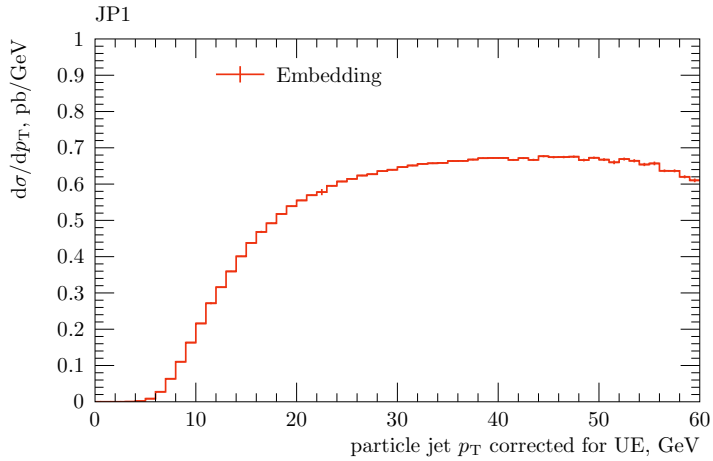
The first one is one minus the (generalized) background rate and the second one is the (generalized) efficiency. The "generalized" qualifier here is meant to convey that they depend not only on the actual detector properties, but also on the specific jet selection cuts, including the cuts applied to implement the binning.

The matching rates are important to monitor when performing the unfolding. The binned rates for a set of wide cuts are shown in Fig. 5.8, which illustrate a few important points about this procedure. The raw background shown in Fig. 5.8a has contributions at low jet p_{T}, presumably mostly from the minbias jets. The raw efficiency shown in Fig. 5.8b has a natural trigger "turn on" behavior at low p_{T} and a slight decrease in efficiency at the highest p_{T} due to the high-p_{T} track cut. Once the analysis binning is set up in a reasonable p_{T} range, the matching rates acquire dips at the new low and high p_{T} due to the jets leaking in and out near those limits. It is possible to tune the binning to expand the detector level cuts to increase efficiency[6], but at the cost of increasing the background. And in the other way, making the particle level cuts more selective can decrease the background, but also decrease the efficiency. In any case, the areas of low efficiency or high background are at risk of higher systematic uncertainties, but if those are thoroughly investigated they should be generally safe. In our case, it is best to avoid areas of extremely low efficiency near the trigger thresholds.

[6]This is sometimes played as an introduction of so-called "garbage bins", which despite the name still require quality data and simulation.



(a) Raw background rate (from Eq. (5.5))



(b) Raw efficiency

Figure 5.8: Raw matching rates calculated for matching of particle jets with p_T from 0 GeV to 60 GeV and $|\eta| < 0.8$ to detector jets of 5-60 GeV and also $|\eta| < 0.8$.

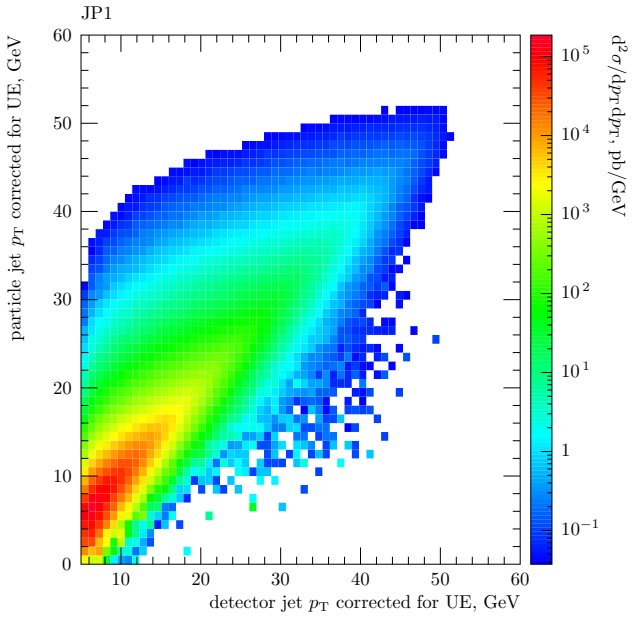


Figure 5.9: Raw covariance between particle and detector spectra of matched jets.

5.5.3 Regularization with the bin size

The takeaway from Section 5.4.2 is that one should not try to bin the data more finely than the detector's resolution capability. In the case of the p_{T}-smearing response, the values of the response matrix should vary sufficiently from bin to bin, so that the statistical uncertainties become small in comparison to the differences in the measured spectrum. A way to achieve this is to choose the p_{T} bin size to be comparable to the p_{T} resolution of the detector. Choosing the right bin size is in itself a form of regularization.

The plot in Fig. 5.10 shows the detector resolution estimated from the finely binned response matrix. This suggests that the p_{T}-bin size should be around 2 GeV at $p_{\mathrm{T}} = 10$ GeV and around 9 GeV at $p_{\mathrm{T}} = 50$ GeV. A more complete scheme for choosing bin sizes can be constructed with the assumption of a linearly growing resolution:

$$\sigma(p_T) = a_0 + a_1 p_T \tag{5.43}$$

for which we get for i-th bin size Δp_{Ti}:

$$\sigma(\Delta p_{T1}) = \Delta p_{T1} \qquad \Rightarrow \qquad \Delta p_{T1} = \frac{a_0}{1 - a_1} \tag{5.44}$$

$$\sigma(2\Delta p_{T1} + \Delta p_{T2}) = \Delta p_{T2} \qquad \Rightarrow \qquad \Delta p_{T2} = \frac{a_0(1 + a_1)}{(1 - a_1)^2} \tag{5.45}$$

$$\sigma(2\Delta p_{T1} + 2\Delta p_{T2} + \Delta p_{T3}) = \Delta p_{T3} \qquad \Rightarrow \qquad \Delta p_{T3} = \frac{a_0(1 + a_1)^2}{(1 - a_1)^3} \tag{5.46}$$

This series of equations can be illustrated by the cartoon in Fig. 5.11. One can recognize that the sequence of bin sizes Δp_{T1}, Δp_{T2}, Δp_{T3}, … is a Geometric series.

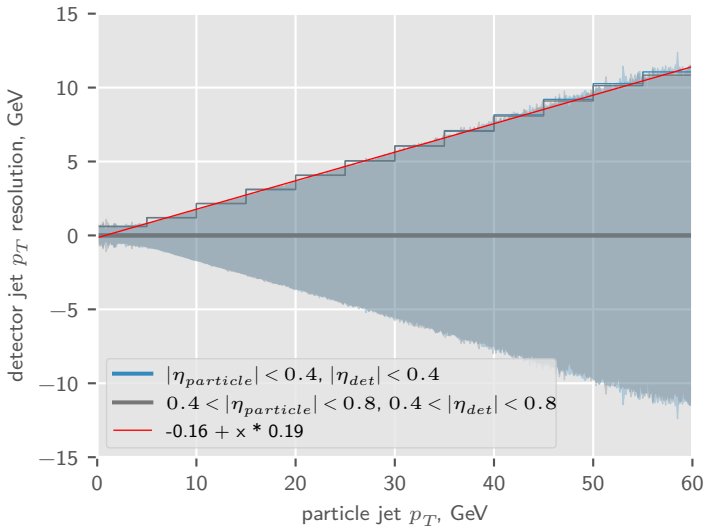


Figure 5.10: Detector resolution estimated from the width of fine response matrix slices. A linear regression fit is overlayed.

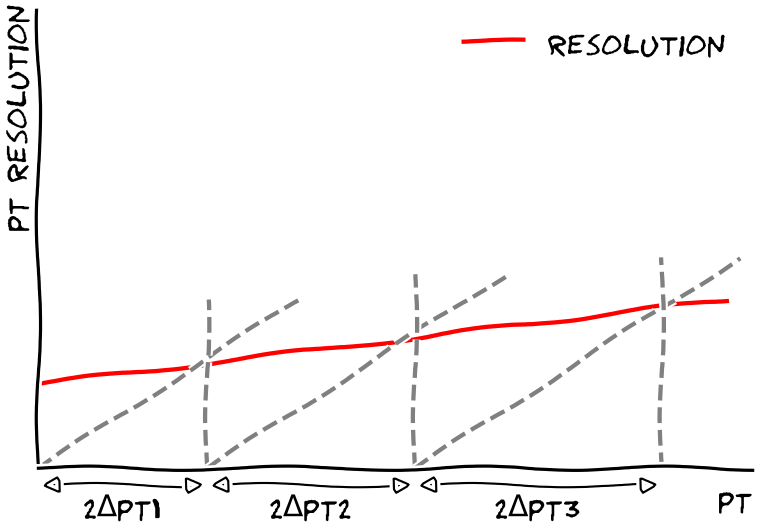


Figure 5.11: Schematic illustration of the solutions (Eqs. (5.44) to (5.46)) for p_{T} bin width Δp_{T} at a given p_{T}.

5.5.4 Unfolding and the Underlying Event correction

As was discussed in Section 1.3.4, the jets in the data have the UE yield estimated on an event-by-event basis. According to this method, the measured detector jet p_{T} scale is effectively adjusted to subtract the event's UE activity momentum density. In order for unfolding to make sense, a similar subtraction needs to be done on the simulation when constructing the response to ensure that its detector level p_{T} scale match the one in the data. If the scale at the detector level is adjusted in the simulation, the subtraction needs to be preserved in the jet p_{T} scale definition at particle level. The correlations between the UE cones of the matching detector and particle jets presented in Fig. 5.12 shows, as expected for particles with a low p_{T}, a good UE resolution for tracking and a decent one from calorimetery.

The UE density discrepancy of 10-20% seen in Fig. 4.17 has the potential to affect the measurement, as the UE-corrected p_{T} spectra shapes may not match between the data and embedding. Ideally this would warrant an extra dimension in the unfolding, or an alternative UE subtraction method, but for this analysis, an additional systematic uncertainty was assigned for this effect in Section 6.5.5.

5.6 Uncertainty due to the simulation statistics

As was shown in Section 5.4.2, the uncertainty analysis is extremely important for doing the unfolding. This analysis must account for not just the uncertainty on the measured spectrum b, but also on the simulation statistics that are available to estimate the detector response.

The individual elements of the response matrix M and the background fraction $\frac{b^{\text{simu.}}_{0;i}}{b^{\text{simu.}}_{i}}$ are formed from the event counts that are subject to statistical fluctuations. The counts of events in the elements of the correlation matrix $A_{ij} = M_{ij}\, x^{\text{simu.}}_{j}$, in elements of the expected truth spectrum $x^{\text{simu.}}_{j}$ and measured spectrum $b^{\text{simu.}}$, and background spectrum $b^{\text{simu.}}_{0;i} = b^{\text{simu.}}_{i} - \sum_j A_{ij}$ are all distributed by a multinomial distribution with the total count fixed by the size of the simulation sample. Propagation of the uncertainty through the operations of the division and, especially, the taking of the matrix inverse M^{-1} is impractical to do analytically, hence a Monte Carlo method is used for this.

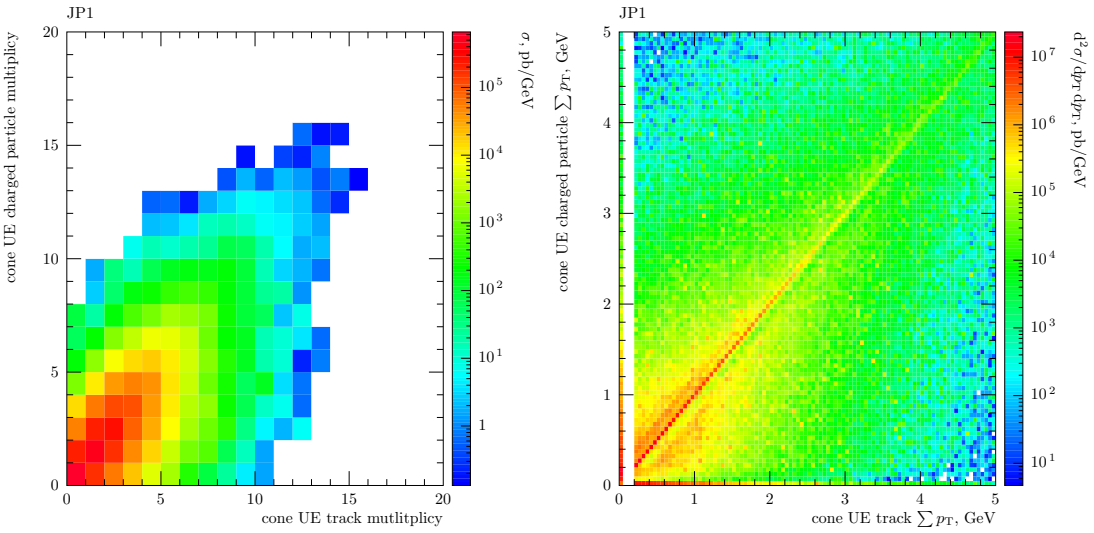
$$A_{ij} \sim \text{Multinomial}(\lambda^{(A)}_{ij}, N) \tag{5.47}$$

$$\left(b_i - \sum_j A_{ij}\right) \sim \text{Multinomial}(\lambda^{(b)}_{i}, N) \tag{5.48}$$

$$\left(x_j - \sum_i A_{ij}\right) \sim \text{Multinomial}(\lambda^{(x)}_{j}, N) \tag{5.49}$$

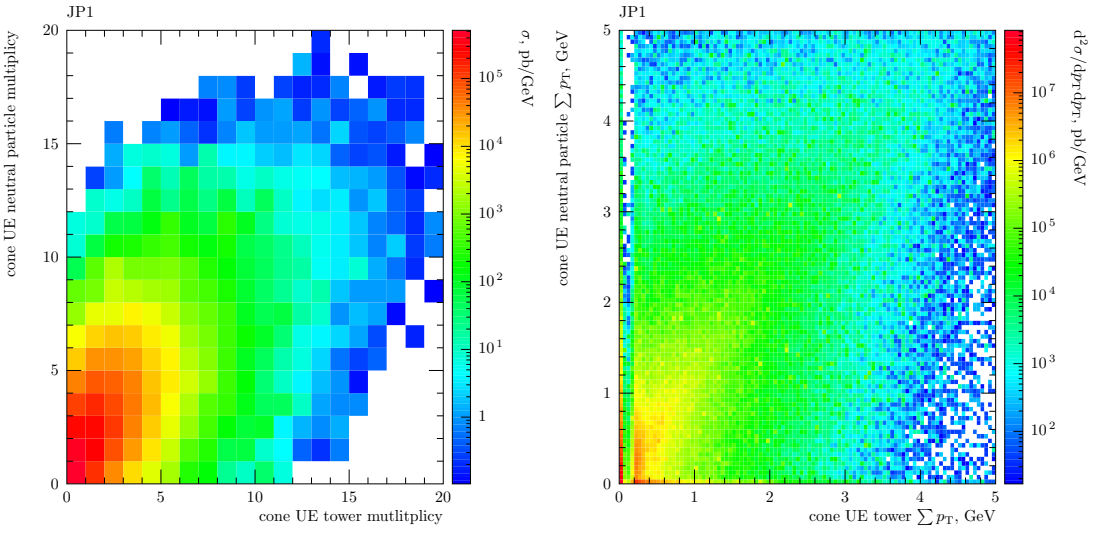
The elements of M were estimated from these sampled replicas (1000 of them), and replica unfolded results were calculated. The standard deviation of the unfolded replicas was taken as the uncertainty.

The first approximation taken here is to replace sampling of the whole spectra from the multinomial distribution with sampling of individual bins from a Poisson distribution. This



(a) Charged particle/Track multiplicity

(b) Charged particle/Track $\sum p_{\mathrm{T}}$



(c) Neutral particle/Tower multiplicity

(d) Neutral particle/Tower $\sum p_{\mathrm{T}}$

Figure 5.12: Multiplicity correlations between contents of the UE cones at the particle and detector levels estimated from the embedding simulation.

resulted in a very similar total number of events, and did not change the result significantly compared to using the multinomial distribution, but was significantly faster to compute.

Technically every entry to the response distribution comes with an independent weight coming from the inverse luminosity corresponding to its partonic p_T bin, and possibly from the soft reweighting or the vertex reweighting. In the first case, the Monte Carlo simulation would sample all the partonic p_T bins separately before merging them with their weights to get the combined response. For the latter cases the following approximation was used:

$$v \simeq \frac{\sum_i w_i \lambda_i^{(v)}}{\sum_i \lambda_i^{(v)}} R^{(j)} \mid R^{(v)} \sim \text{Poisson}\left(\sum_i \lambda_i^{(v)}\right) \tag{5.50}$$

for v being one of the distributions that has contributions with different weights:

$$v = \sum_i w_i R_i^{(v)} \mid \text{For all } i\text{: } R_i^{(v)} \sim \text{Poisson}(\lambda_i^{(v)}) \tag{5.51}$$

Another limitation of this method is that values of $\lambda_i^{(j)}$ and all other necessary combinations were estimated from our single simulation sample. Thus, bins that didn't have any entries in them will remain empty in all of the replicas. Bins that didn't have many statistics in the first place will also not yield a precise estimate of $\lambda_i^{(j)}$.

Chapter 6

Analysis and results

At its core, the present cross section measurement is a simple counting experiment with a correction for detector effects, detector and trigger live time, trigger efficiency, trigger prescalers and normalization for the collected luminosity. In this chapter the insights from the previous chapters will be combined into a comprehensive description of the measurement procedure.

6.1 Detector level cross section

To begin, here is a definition of the detector level inclusive jet cross section (denoted with "b" instead of "σ" to match the notation from Chapter 5):

$$b_i^{\text{data}} \equiv \sum_{\text{events}} \Theta(\text{event cuts})\, \omega_{\text{event}}^{\text{data}} \times \\ \times \sum_{\substack{\text{detector level} \\ \text{jets in event}}} \Theta(\text{detector jet cuts})\, \Theta((\eta^{\text{jet}}, p_{\text{T}}^{\text{jet}} - \Delta p_{\text{T}}^{\text{UE}}) \in \text{bin } i) \quad (6.1)$$

where the "event cuts" are given in Table 6.1 and the "detector jet cuts" are those listed in Table 3.4.

Cut
• Must have a vertex with a positive (non-pileup) rank and $\lvert v_z \rvert < 60$ cm
• JPi fired (not applicable to the simulation)
• JPi should fire

Table 6.1: Event level cuts used at the detector level for selecting Jet Patch triggered events.

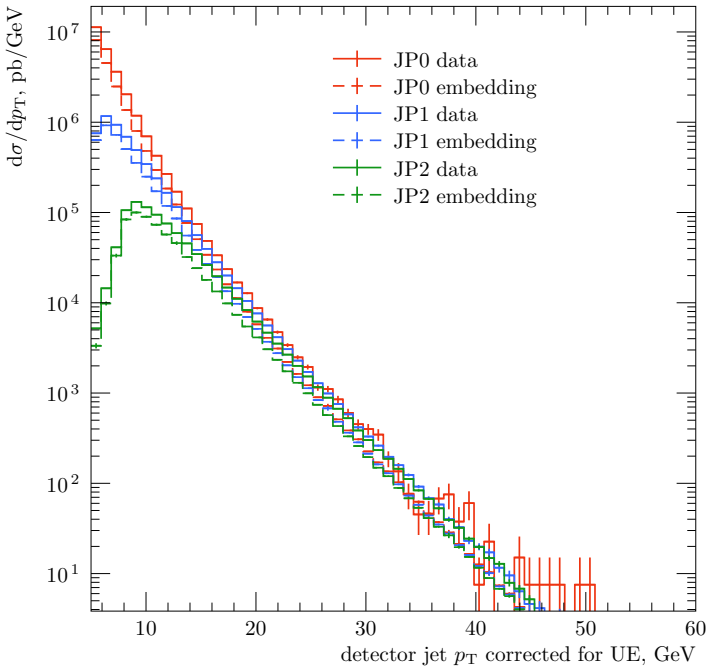


Figure 6.1: Differential cross section for inclusive jet production at detector level for different Jet Patch triggers.

Each run in the data has a different luminosity associated with it, hence, for a physical process for which N_i events were detected with overall efficiency ϵ_i in the run with index i, its measured cross section will be[1]:

$$\sigma_r = \frac{N_r}{\epsilon_r \, \mathcal{L}_r} \tag{6.2}$$

Now, to combine the results for this intensive observable among several runs one has to choose how they will be weighted. The choice taken here is to minimize the dominant statistical uncertainty on the N_r events of a jet production process that is rare relative to the ZDC rate used to monitor $\mathcal{L}_r$:

$$\delta\{\sigma_r\} \sim \sqrt{N_r} \sim \sqrt{\epsilon_r \, \mathcal{L}_r} \tag{6.3}$$

Using the result from Appendix B, and using the weight from Eq. (B.6) one gets the following averaging:

$$\langle\sigma\rangle = \frac{\sum_r \sigma_r \frac{\epsilon_r \, \mathcal{L}_r}{\sum_j \epsilon_j \, \mathcal{L}_j}}{\sum_r \frac{\epsilon_r \, \mathcal{L}_r}{\sum_j \epsilon_j \, \mathcal{L}_j}} = \frac{\sum_r N_r}{\sum_r \epsilon_r \, \mathcal{L}_r} \tag{6.4}$$

Hence each event within a run is weighted with the same weight:

$$\omega_{\text{event}}^{\text{data}} = \frac{1}{\sum_r \epsilon_r \, \mathcal{L}_r} \tag{6.5}$$

[1]Formally, the σ is not dependent on the run id i, but in a measurement, of course, it will fluctuate.

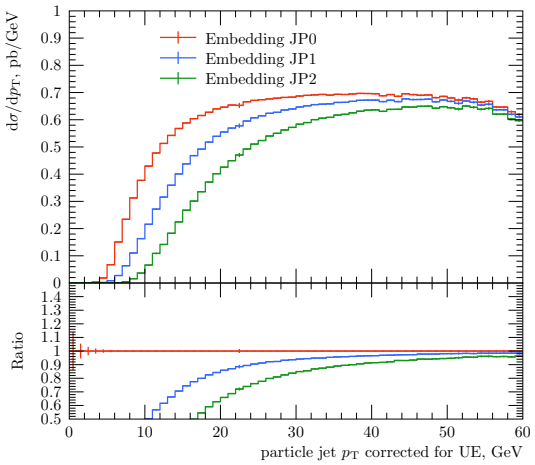


Figure 6.2: Raw generalized efficiency calculated for matching of particle jets with p_{T} from 0 GeV to 60 GeV and $|\eta| < 0.8$ to detector jets of 5-60 GeV and also $|\eta| < 0.8$.

which matches a commonly known fact that non-uniform weighting of events causes larger relative statistical uncertainties.

Writing down all run-dependent efficiencies that are not part of the unfolding one gets:

$$\epsilon_r = \frac{1}{\mathrm{PS}(\mathrm{JP}i)_r} \times (\text{Detector live time fraction})_r \times (\text{Trigger live time fraction})_r \qquad (6.6)$$

There are two live times to consider here: the trigger live time fraction is governed by trigger time outs internal to the TCU, whereas the detector live time fraction is reported to the trigger system based on the detector state.

The cross sections at detector level observed in the data and simulation are shown in Fig. 6.1. The cross sections seen by different triggers agree at jet p_{T} values that are much above the trigger thresholds. In general, they agree reasonably well with the embedding simulation.

6.2 Combining different triggers

Among the three available jet patch triggers there is the JP0 that has a higher jet detection efficiency at low $p_{\mathrm{T}}^{\mathrm{jet}}$ (see Fig. 6.2), but poor statistics at high $p_{\mathrm{T}}^{\mathrm{jet}}$ due to a large prescale, and on the other end of the spectrum, JP2 has two orders of magnitude more of statistics at high $p_{\mathrm{T}}^{\mathrm{jet}}$, but a lower efficiency at low $p_{\mathrm{T}}^{\mathrm{jet}}$ due to the higher threshold. Naturally, it is desirable to use the data from all available triggers to gain the possible benefits for different regions of phase space.

A complication arises with regard to using the simulation to unfold distributions with mixed triggers. The problem arises with the treatment of different prescales. A brute force approach of throwing away events from the simulation according to the data prescale factors would lead to a waste of simulation's statistics, so instead the prescales are applied in reweighting. Other problems arise with the treatment of the overlaps between the triggers.

Category	Condition in data
JP0 promoted	JP0 fired $\wedge$ $\overline{\text{JP1 should fire}}$ $\wedge$ JP0 should fire
JP1 promoted	(JP0 fired $\vee$ JP1 fired) $\wedge$ $\overline{\text{JP2 should fire}}$ $\wedge$ JP1 should fire
JP2 promoted	(JP0 fired $\vee$ JP1 fired $\vee$ JP2 fired) $\wedge$ JP2 should fire
JP0 demoted	JP0 fired $\wedge$ JP0 should fire
JP1 demoted	$\overline{\text{JP0 fired}}$ $\wedge$ JP1 fired $\wedge$ JP1 should fire
JP2 demoted	$\overline{\text{JP0 fired}}$ $\wedge$ $\overline{\text{JP1 fired}}$ $\wedge$ JP2 fired $\wedge$ JP2 should fire

Table 6.2: Definitions of the conditions used to combine different triggers in the data.

Category	Condition in simulation	Weight in simulation
JP0 promoted	$\overline{\text{JP1 should fire}}$ $\wedge$ JP0 should fire	$\frac{1}{\text{PS(JP0)}}$
JP1 promoted	$\overline{\text{JP2 should fire}}$ $\wedge$ JP1 should fire	$\left(\frac{1}{\text{PS(JP0)}} + \frac{1}{\text{PS(JP1)}} - \frac{1}{\text{PS(JP0)PS(JP1)}}\right)$
JP2 promoted	JP2 should fire	1
JP0 demoted	JP0 should fire	$\frac{1}{\text{PS(JP0)}}$
JP1 demoted	JP1 should fire	$\frac{1}{\text{PS(JP1)}}\left(1 - \frac{1}{\text{PS(JP0)}}\right)$
JP2 demoted	JP2 should fire	$\left(1 - \frac{1}{\text{PS(JP0)}}\right)\left(1 - \frac{1}{\text{PS(JP1)}}\right)$

Table 6.3: Definitions of the conditions used to combine and reweight different triggers in the simulation.

For a given trigger, the relative fraction of its overlap with other triggers depends on its rate and prescale factor, as well as the rates and prescale factors of the other triggers that can overlap with it. In the simulation sample, the higher threshold jet patch triggers are necessarily a subset of the lower threshold jet patch triggers, so the overlap fractions don't match those found the data. Another issue with overlaps is that they lead to double counting when the trigger samples are simply added together.

The resolution to this complexity lies in getting rid of overlaps by defining event categories different from JP0, JP1, JP2, that would not overlap in the data. Then the simulation sample can be reweighted to match the data. This gives rise to two kinds of schemes for doing this: trigger promotion and trigger demotion. The definitions of categories are given in Tables 6.2 and 6.3 and the derivation of weights can be found in Appendix C. The cuts listed in that tables are used in place of the "JPi should fire" in Table 6.1. And the factor of $\frac{1}{\text{PS(JP}i\text{)}}$ is also removed from Eq. (6.6); instead the weight from the table is applied to the simulation when constructing the detector response. The three categories in the data can then be treated as a single sample, and the same is done for the three embedding categories with the respective weights applied.

The trigger promotion differs from demotion in that it defines non-overlapping embedding categories. This prevents one from overestimating the statistical precision of the simulation sample when these categories are combined. On the other hand, trigger demotion conservatively uses all available simulation statistics. As seen in Fig. 6.3, both methods give the same

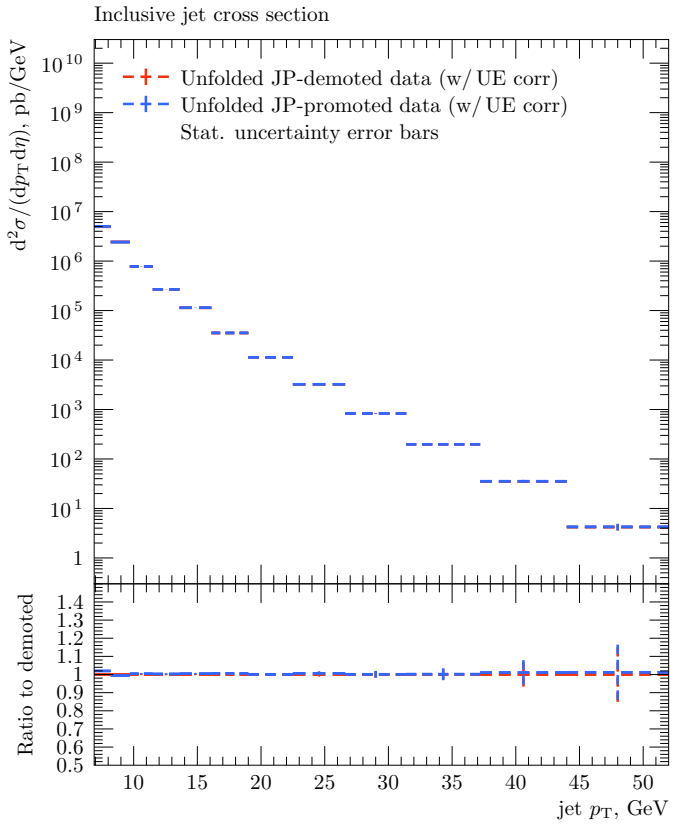


Figure 6.3: Comparison of spectra unfolded using the two different trigger combination schemes described in Tables 6.2 and 6.3.

$\lvert\eta\rvert$ bin	Detector level jet p_{T} bin boundaries, GeV
$0.0 < \lvert\eta\rvert < 0.4$	6.9, 8.2, 9.7, 11.5, 13.6, 16.1, 19.0, 22.5, 26.6, 31.4, 37.2, 44.0, 52.0
$0.4 < \lvert\eta\rvert < 0.8$	6.9, 8.2, 9.7, 11.5, 13.6, 16.1, 19.0, 22.5, 26.6, 31.4, 37.2, 44.0, 52.0
$\lvert\eta\rvert$ bin	Particle level jet p_{T} bin boundaries, GeV
$0.0 < \lvert\eta\rvert < 0.4$	6.9, 8.2, 9.7, 11.5, 13.6, 16.1, 19.0, 22.5, 26.6, 31.4, 37.2, 44.0, 52.0
$0.4 < \lvert\eta\rvert < 0.8$	6.9, 8.2, 9.7, 11.5, 13.6, 16.1, 19.0, 22.5, 26.6, 31.4, 37.2, 44.0, 52.0

Table 6.4: List of all 2D bins in the jet η-p_{T} space in the binning of the detector and particle level cross sections.

cross section.

6.3 Unfolding the detector response

The particle level jet cross sections x_j^{data} are obtained from the measured detector level cross sections b_i^{data} according to the following relation:

$$x_j^{\mathrm{data}} = \sum_i \mathcal{M}_{ij}^{-1}\, b_i^{\mathrm{data}} \tag{6.7}$$

where the response matrix $\mathcal{M}$ is given by

$$\mathcal{M}_{ij} = \left(\frac{b_i^{\mathrm{simu.}}}{\sum_k A_{ik}}\right)\left(\frac{A_{ij}}{x_j^{\mathrm{simu.}}}\right) \tag{6.8}$$

Three components are required to construct the response matrix: the particle level spectrum $x_j^{\mathrm{simu.}}$, the simulation's detector level spectrum $b^{\mathrm{simu.}}$ and the covariance of particle and detector spectra of matched jets A_{ij}.

$$x_j^{\mathrm{simu.}} \equiv \sum_{\text{events}} \omega_{\text{event}}^{\mathrm{simu.}} \sum_{\substack{\text{particle level}\\ \text{jets in event}}} \Theta((\eta^{\mathrm{jet}}, p_{\mathrm{T}}^{\mathrm{jet}} - \Delta p_{\mathrm{T}}^{\mathrm{UE}}) \in \text{bin } j) \tag{6.9}$$

There are no particle level cuts other than the selection for specific η-p_{T} bins. The weight $\omega_{\text{event}}^{\mathrm{simu.}}$ is a product of the partonic p_{T} bin weights (Section 4.1.2) with the fudge factors (Section 4.1.3), the soft reweighting weight (Section 4.1.4), and the vertex weight (Section 4.2.1):

$$\omega_{\text{event}}^{\mathrm{simu.}} \equiv \frac{\sigma_{\hat{p}_{\mathrm{T}}\text{ bin}}}{N_{\hat{p}_{\mathrm{T}}\text{ bin}}}\, f_{\text{partonic. bin}}\, \omega(\hat{p}_{\mathrm{T}})\, \omega(v_z) \tag{6.10}$$

The absolute magnitude of the partonic bin cross section does not affect the unfolding procedure, it is only needed to get the right weight relative to the other partonic bins.

The definition of $b^{\mathrm{simu.}}$ is nearly identical to the one for $b^{\mathrm{data.}}$ (Eq. (6.1)):

$$b_i^{\mathrm{simu.}} \equiv \sum_{\text{events}} \Theta(\text{event cuts})\omega_{\text{event}}^{\mathrm{simu.}} \sum_{\substack{\text{detector level}\\ \text{jets in event}}} \Theta(\text{detector jet cuts})\Theta((\eta^{\mathrm{jet}}, p_{\mathrm{T}}^{\mathrm{jet}} - \Delta p_{\mathrm{T}}^{\mathrm{UE}}) \in \text{bin } i) \tag{6.11}$$

where, just as for the data, the "event cuts" are listed in Table 6.1 and the "detector jet cuts" are provided in Table 3.4.

$$A_{ij} \equiv \sum_{\text{events}} \Theta(\text{event cuts})\, \omega_{\text{event}}^{\text{simu.}} \times \\ \times \sum_{\substack{\text{pairs of} \\ \text{matched jets}}} \Theta((\eta^{\text{detector jet}}, p_{\text{T}}^{\text{detector jet}} - \Delta p_{\text{T}}^{\text{UE}}) \in \text{bin } i)\, \Theta(\text{detector jet cuts}) \times \\ \times \Theta((\eta^{\text{particle jet}}, p_{\text{T}}^{\text{particle jet}} - \Delta p_{\text{T}}^{\text{UE}}) \in \text{bin } j)) \quad (6.12)$$

where the matched jet pairs are defined according to the algorithm defined in Section 5.5.1.

The boundaries for the η-p_{T} bins used in the definition above are provided in Table 6.4. A total of 26 bins are used at both particle and detector levels, which means that the dimension of x and b is 26 and the dimension of A is 26×26. Bin boundaries at the particle level are chosen to form a geometric series according to Section 5.5.3, assuming that the detector bins are identical.[2]

6.4 Data statistics uncertainty

The statistics available in the data are limited, especially for bins at high p_{T}. The statistical uncertainty of the detector level spectrum can be quantified using the covariance matrix for b:

$$B \equiv \text{cov}(b, b^T) \quad (5.32 \text{ revisited})$$

Technically, it should contain non-diagonal elements coming from the events that had two or more jets contributing to the inclusive spectrum, but for this analysis this detail was ignored, and a simple approximation was used:

$$B = \text{diag}(\{b_i\}) \quad (6.14)$$

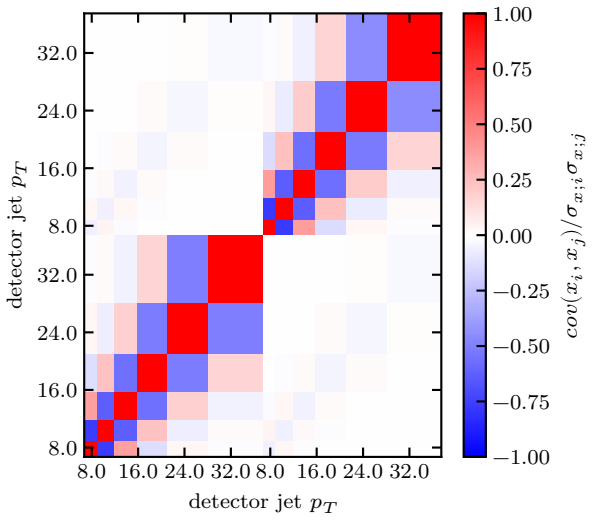
The covariance matrix for x is definitely not diagonal, as there are significant correlations between bins. Substituting Eq. (6.7) into Eq. (5.32) one gets:

$$X = \text{cov}(x, x^T) = M^{-1} B (M^{-1})^{\text{T}} \quad (6.15)$$

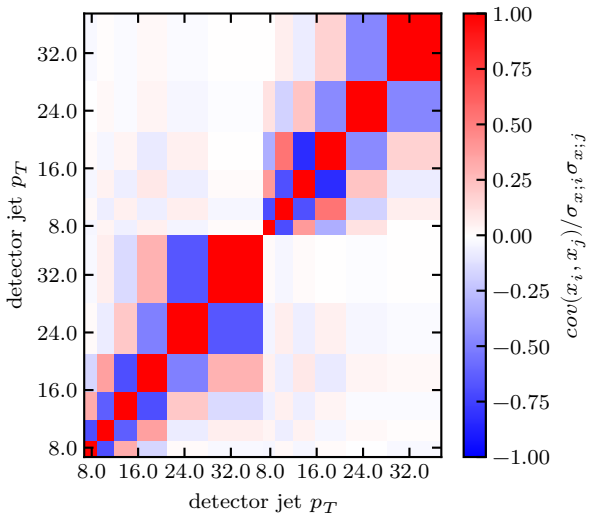
The magnitudes of the statistical uncertainties in the data are tightly coupled with the choice of p_{T} binning.

The general structure (without the magnitude) of the correlation between the different bins of the unfolded cross section is visualized in Fig. 6.4a. Naturally, it inherits the same "oscillating" behavior as M^{-1}.

[2] This is not the optimal configuration. In fact, the actual implementation divides the particle jet p_{T} into 600 bins from 0 to 60 GeV and detector jet p_{T} into 550 bins from 5 to 55 GeV. The histograms are then rebinned to the final binnning before constructing the response. This should, in principle, allow one to do a numeric optimization for the unfolding binning using, e.g., annealing techniques.



(a) due to the data statistics



(b) due to the simulation statistics

Figure 6.4: Pearson correlation coefficient $\mathrm{cov}(x_i, x_j)/\sqrt{\mathrm{cov}(x_i, x_i)\,\mathrm{cov}(x_j, x_j)}$ for the JP0 trigger.

6.5 Systematic uncertainties

6.5.1 Embedding simulation sample statistics

The approach to determining this systematic uncertainty is given in Section 5.6. Just as for the data, this uncertainty is also tied to the p_{T} binning. The number of bins used puts even higher pressure on the simulation statistics since the number of elements in the response matrix grows quadratically with the number of bins.

The Pearson correlation coefficient for fluctuations due to the simulation statistics is shown in Fig. 6.4b.

6.5.2 Jet Energy Scale

One of the most important quantites for a jet is its transverse momentum p_{T}. Accuracy of the transverse momentum determination is essential to the measurement of the differential jet cross section. This can be illustrated with a simple example of a cross section that has a power-law dependence on the momentum:

$$\frac{d\sigma}{dp_T} \simeq A\, p_T^{-\alpha} \tag{6.16}$$

with α on the order of 4-6. Assuming that neutral energy is measured precisely, and in the limit of large track multiplicity and precise measurement of the track p_{T}, the cross section at detector level will be

$$\frac{d\sigma_{\mathrm{det}}}{dp_T} \simeq A(\epsilon(1-R_T)+R_T)^{-\alpha} p_T^{-\alpha} \tag{6.17}$$

where ϵ is the efficiency, and R_T is the neutral energy fraction.

The relative uncertainty on this detector level cross section is then given by:

$$\frac{\delta\left\{\frac{d\sigma_{\mathrm{det}}}{dp_T}\right\}}{\frac{d\sigma_{\mathrm{det}}}{dp_T}} = \frac{\alpha}{\epsilon(1-R_{\mathrm{T}})+R_{\mathrm{T}}}\left(\epsilon(1-R_{\mathrm{T}})\,\frac{\delta\{\epsilon\}}{\epsilon}+\right.$$
$$\left.+(1-R_{\mathrm{T}})p_{\mathrm{T}}\frac{\delta\{(1-R_{\mathrm{T}})p_{\mathrm{T}}\}}{(1-R_{\mathrm{T}})p_{\mathrm{T}}}+\epsilon R_{\mathrm{T}}p_{\mathrm{T}}\,\frac{\delta\{R_{\mathrm{T}}p_{\mathrm{T}}\}}{R_{\mathrm{T}}p_{\mathrm{T}}}\right) \tag{6.18}$$

Assuming $\epsilon = 90\%$ and $R_{\mathrm{T}} = \frac{1}{3}$, this evaluates to

$$\frac{\delta\left\{\frac{d\sigma_{\mathrm{det}}}{dp_T}\right\}}{\frac{d\sigma_{\mathrm{det}}}{dp_T}} = \alpha\left(\frac{9}{14}\,\frac{\delta\{\epsilon\}}{\epsilon}+\frac{9}{14}\,\frac{\delta\{(1-R_{\mathrm{T}})p_{\mathrm{T}}\}}{(1-R_{\mathrm{T}})p_{\mathrm{T}}}+\frac{5}{14}\,\frac{\delta\{R_{\mathrm{T}}p_{\mathrm{T}}\}}{R_{\mathrm{T}}p_{\mathrm{T}}}\right) \tag{6.19}$$

This shows how a Jet Energy Scale (JES) uncertainty associated with the momentum of charged particles is contributing an approximately twice larger relative uncertainty to the jet p_{T} and to the detector level cross section uncertainty than the momentum of neutral particles. This behavior is illustrated by Fig. 6.5 showing an estimation for effect of a 4% benchmark uncertainty on the jet p_{T} distribution.

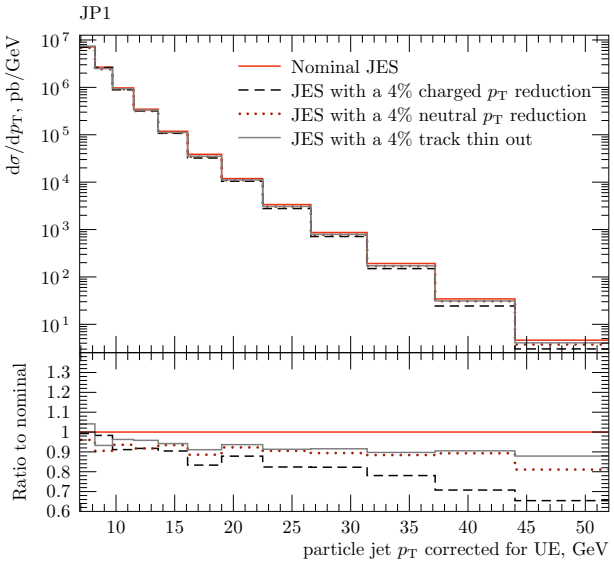


Figure 6.5: Jet p_{T} spectra obtained using a nominal JES, or a JES where neutral or charged energy is scaled down by 4% or with a JES for which 4% of tracks are removed from each event randomly.

The approach used in this work to estimate the Jet Energy Scale systematic uncertainties is to process the simulation sample with each jet's p_{T} shifted up or down according to a given systematic effect. After performing the unfolding of the data with those modified simulation samples, the resulting cross sections are compared to the nominal cross section and the difference is taken as the uncertainty. The same shifts can be equivalently applied to the data, and in all considered cases this yields nearly identical uncertainties to the shifts in simulation.

EMC calibration and response to hadrons

A standard formula for evaluating the uncertainty on the contribution of the BEMC to jet p_{T} that was used for previous jet spin asymmetry analyses in STAR looks like (proof given in Appendix E):

$$\frac{\delta\{p_{\mathrm{T}}^{\mathrm{jet}}\}}{p_{\mathrm{T}}^{\mathrm{jet}}} = \sqrt{R_{\mathrm{T}}^2 \left(\frac{\delta\{E_{\mathrm{T}}^{\mathrm{tower}}\}}{E_{\mathrm{T}}^{\mathrm{tower}}}\right)^2 + (1-R_{\mathrm{T}})^2 \left(\frac{\delta\{E_{\mathrm{T}}^{\mathrm{had.}}\}}{E_{\mathrm{T}}^{\mathrm{had.}}}\right)^2} \tag{6.20}$$

where the first contribution

$$\frac{\delta\{E_{\mathrm{T}}^{\mathrm{tower}}\}}{E_{\mathrm{T}}^{\mathrm{tower}}} \equiv \sqrt{\left(\frac{\delta\{\mathrm{gain}\}}{\mathrm{gain}}\right)^2 + \left(\frac{\delta\{\text{tower eff.}\}}{\text{tower eff.}}\right)^2} \tag{6.21}$$

scales with neutral energy and represents the uncertainty related to the calibration of the calorimeter's response to the EM-showers (that, for measurement of jets, are caused by γ's from π^0 decays).

Source	Uncertainty
inner BEMC ring $\frac{\delta\{\text{gain}\}}{\text{gain}}$	3.0% [42]
outer BEMC ring $\frac{\delta\{\text{gain}\}}{\text{gain}}$	3.8% [42]
$\frac{\delta\{\text{tower eff.}\}}{\text{tower eff.}}$	1%
ϵ_{trk}	90%
S_{hadron}	1.16 [59] (can be estimated from Fig. 6.9 and Fig. 6.10)
$f_{\text{had. resp.}}$	32% [68]
$\frac{\delta\{f_{\text{had. resp.}}\}}{f_{\text{had. resp.}}}$	6% [68]
p_{proj}	72% [68]

Table 6.5: Parameters used in the calculation of the EMC contribution to the JES uncertainty.

The second contribution

$$\frac{\delta\{E_{\text{T}}^{\text{had.}}\}}{E_{\text{T}}^{\text{had.}}} \equiv \left(\frac{1}{\epsilon_{\text{trk}}} S_{\text{hadron}} - p_{\text{proj}}\right) \times f_{\text{had. resp.}} \times \frac{\delta\{f_{\text{had. resp.}}\}}{f_{\text{had. resp.}}} \tag{6.22}$$

scales with the charged transverse momentum fraction and corresponds to the uncertainty due to the response to hadronic showers. The formalism looks a bit more complicated this time, as it accounts for the hadronic subtraction procedure (described in Section 3.4.4) which partially reduces the uncertainty as indicated by subtraction of p_{proj}[3].

The EEMC uncertainties are conservatively estimated to be about the same as the BEMC.

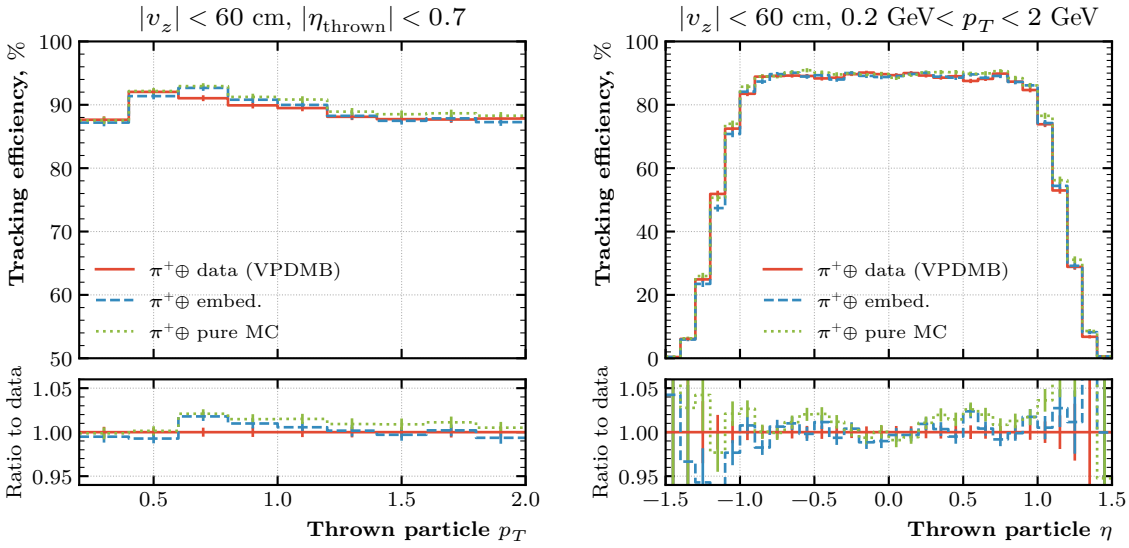
The values of the parameters for Eq. (6.21) and Eq. (6.22) are given in Table 6.5. The uncertainty for the 4 "outer" BEMC rings can be neglected as the majority of 36 rings are associated with the "inner" BEMC ring uncertainty. This evaluates to $\frac{\delta\{E_{\text{T}}^{\text{tower}}\}_{\text{calib.}}}{E_{\text{T}}^{\text{tower}}} = 3.2\%$ and $\frac{\delta\{E_{\text{T}}^{\text{had.}}\}}{E_{\text{T}}^{\text{had.}}} = 1.1\%$.

From Table 3.1 one can infer the value of a conversion factor from GeV to DSM ADC units which is ≈ 4.2 (DSM ADC units)/GeV. A 3% EMC energy uncertainty for an average trigger threshold thus corresponds to about 1 DSM ADC unit. This uncertainty can be estimated by adjusting the Jet Patch trigger thresholds in the trigger simulator.

TPC tracking efficiency

The original study of the TPC tracking efficiency was done in [78] using 2006 data and simulations for pp collisions with jet production at $\sqrt{s} = 200$ GeV. In that study a discrepancy was found in the tracking efficiency between simulated π^+ and π^- tracks embedded into collision data and simulated tracks in a simulation sample. The study finds a 5.9% discrepancy and, after making some assumptions, concludes with a 3.3% tracking efficiency uncertainty for mid-rapidity tracks. Another relevant observation made in that study was that the tracking efficiency appears to be consistent for tracks inside the jet cones and those outside.

[3] Moreover, a better hadronic subtraction procedure that could subtract energy from clusters of towers instead of individual towers could increase the value of p_{proj} to be even closer to 1.



(a) function of track p_{T}

(b) function of track η

Figure 6.6: Tracking efficiency for simulated tracks as a function of track p_T and track η. Simulated tracks are thrown in the context of the data, embedding simulation and simulation without the embedding (labeled "pure MC")

.

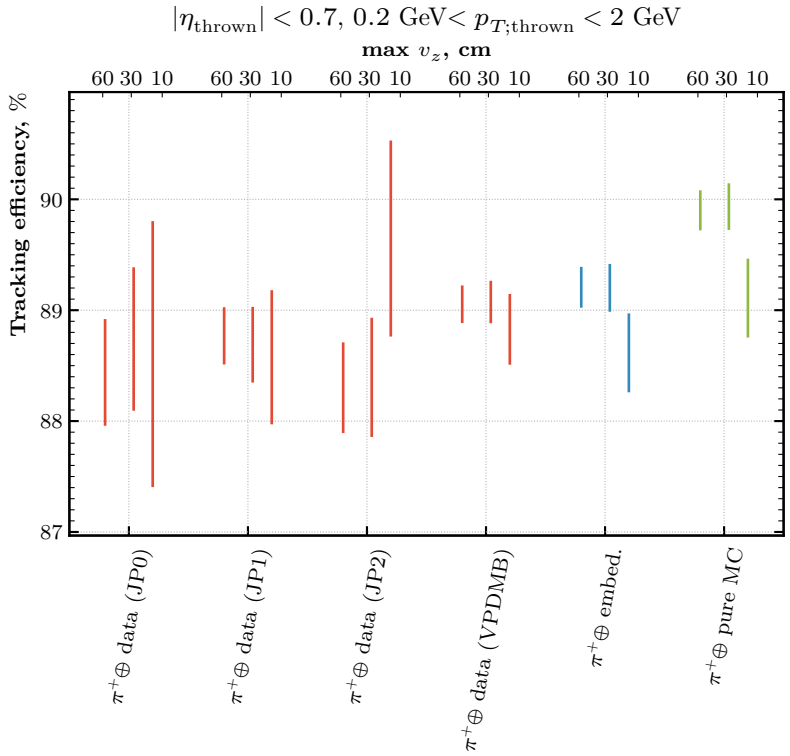
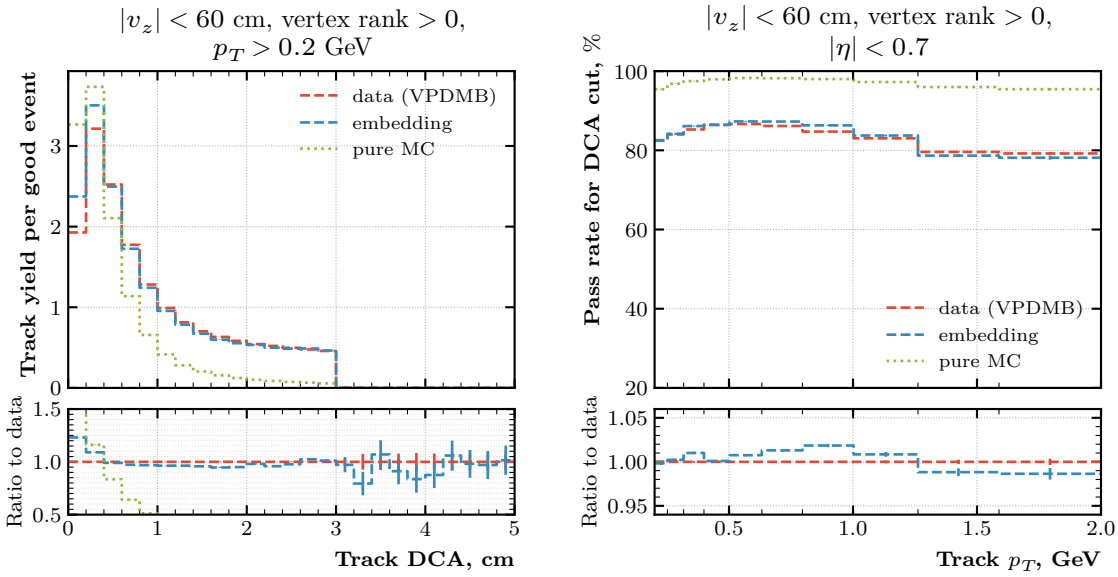


Figure 6.7: Tracking efficiency for simulated tracks in the mid-rapidity region for various values of vertex z cut. Simulated tracks are thrown in the context of the data, embedding simulation and simulation without the embedding (labeled "pure MC").



(a) Distribution of the track DCA

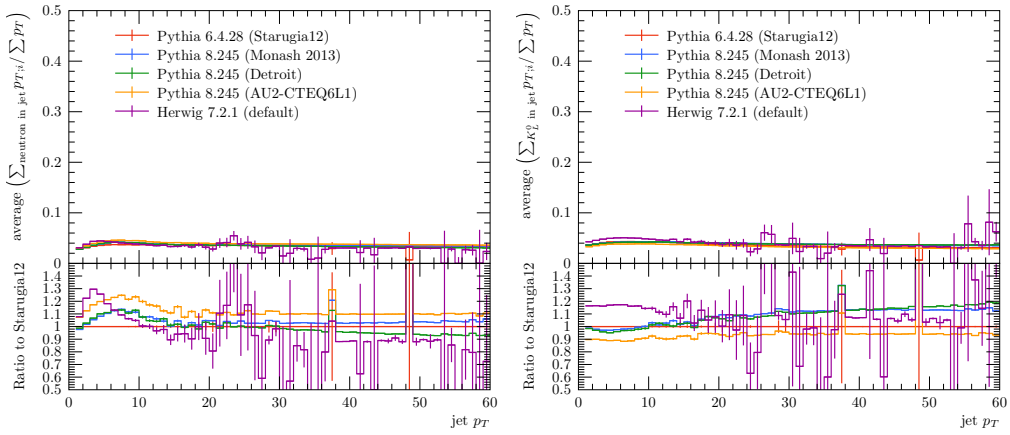
(b) Cut pass rate for track DCA cut

Figure 6.8: Panel (a) shows the distribution of the track DCA, which is related to the cut pass rate for the track DCA cut shown in panel (b).

As a part of the present measurement, an attempt to reevaluate and, if possible, reduce this uncertainty was made. A full technical description of this study is provided in [70] and only a summary of it is given here. Firstly, the present study was redone using the 2012 data and simulation, featuring many improvements, including a new TPC DAQ and updated versions of the reconstruction and simulation software. A few changes were made to improve the previous procedure: for the simulation, the simulated events were embedded in zero bias data, which had not been done previously; and a more consistent procedure was used for selecting the vertex position for the introduced simulated π track, which was placed at the reconstructed vertex of the original event in the simulation, the same way it was done for the data. Figure 6.6 shows the general shape of the TPC tracking efficiency. Single-number summaries of the tracking efficiencies for the simulated tracks are presented in Fig. 6.7, showing a sub-percent agreement between the data and the embedding simulation.

In order to evaluate the new tracking efficiency systematic uncertainty, the fraction of tracks that pass cuts listed in Table 3.2 were evaluated for tracks in the data and the embedding simulation. Most distributions show excellent agreement except for the track DCA which has a noticeable systematic difference as seen in Fig. 6.8a, which likely has to do with an incomplete simulation of the TPC distortion effects. The cut pass rates for DCA shown in Fig. 6.8b have about a 1% discrepancy between the data and embedding.

As seen in Fig. 6.5, the 1% tracking efficiency uncertainty is not equivalent to a 1% charged shift for the jet p_{T}, but only achieves half of the expected effect on the jet p_{T} spectra. Together with the small magnitude of the 1% number, this results in an uncertainty on the jet cross section that is negligible in comparison to the uncertainty due to the EMC response.



(a) neutrons (b) K_0^L's

Figure 6.9: The average fraction of the jet p_{T} carried by long-living neutral hadrons plotted as a function of jet p_{T}. Different lines represent predictions from different Event Generators and their tunes. The "Starugia12" tune refers to the tune described in Section 4.1.1.

TPC calibration

The TPC calibration uncertainty is estimated to be below 1%. This number is based on reconstructed mass peaks of weak decays of Λ and K_{s} matching their known masses within 0.3%.

Hadron fractions

A question can arise whether absence of a hadron calorimeter is detrimental to STAR's ability to do precision jet measurements. The argument can be made that the energy carried by long-lived neutral hadrons is not constrained at all in a jet measurement performed at STAR. Figure 6.9 shows that the fraction of the jet p_{T} carried by such particles is about 4% each, and vastly different hadronization models predict a variation in the fraction of up to about 10%, which together correspond to roughly a 0.4% shift in the jet p_T.

6.5.3 R_{T} discrepancy

The discrepancy in R_{T} distributions seen in Fig. 4.14 may be indicative of a possibly overlooked effect. Different event generator models predict identical values of R_{T} at the particle level (Fig. 6.10), so we do not expect that what is observed is an effect from the underlying physics of jet production. This indicates the presence of a detector effect.

In order to estimate an uncertainty associated with this discrepancy, the simulation sample was reweighted to match the $\mathrm{d}P/\mathrm{d}R_{\mathrm{T}}$ distribution from the data within some slices in the jet p_{T}. The resulting shift (shown in Fig. 6.11) was taken as the uncertainty due to the observed R_{T} discrepancy.

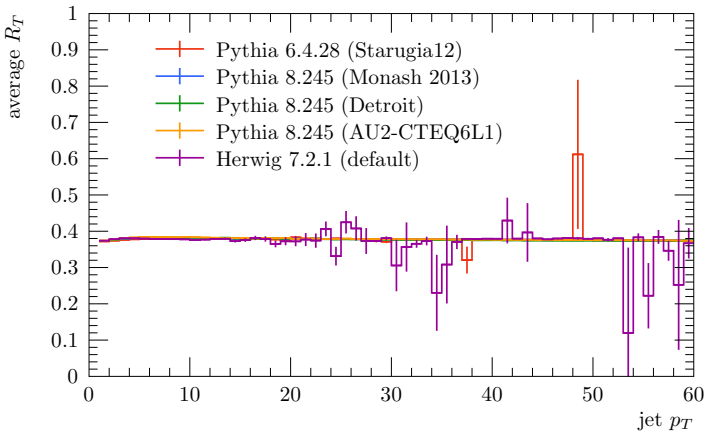


Figure 6.10: Average jet R_{T} plotted as a function of jet p_{T}. Different lines represent predictions from different Event Generators and their tunes. The "Starugia12" tune refers to the tune described in Section 4.1.1.

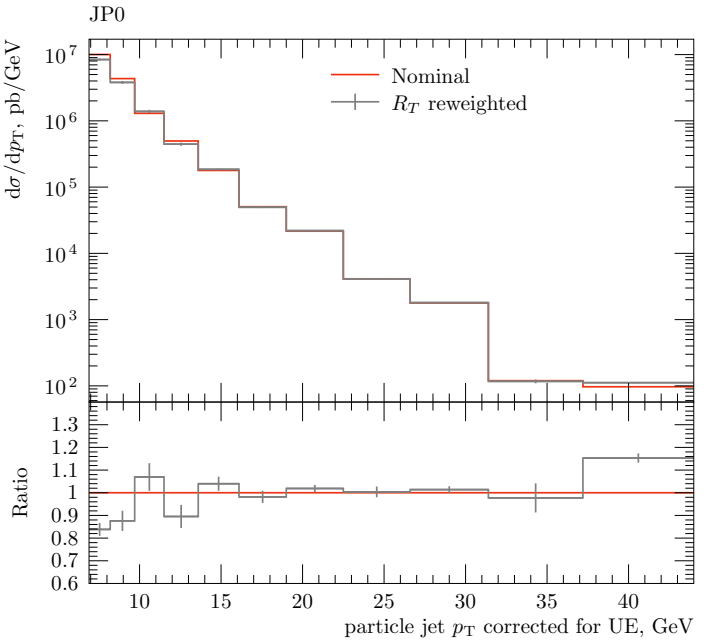


Figure 6.11: Unfolded jet p_{T} spectra using a nominal embedding simulation and one using an embedding simulation reweighted to match the R_{T} spectra of jets at multiple different jet p_{T}. The error bars represent uncertainties due to the limited embedding simulation sample size.

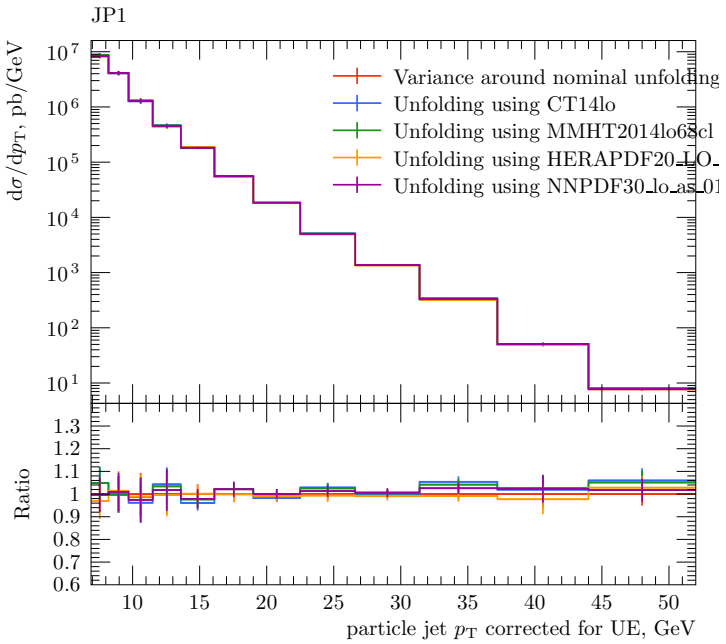


Figure 6.12: Inclusive jet cross sections unfolded from data using simulations that are reweighted with different PDF sets.

6.5.4 Variation of the choice of Parton Distribution Functions used for embedding simulation

The detector response matrix is constructed from the prior given by PYTHIA and its specific tune and the associated bin migration predicted by the detector simulation based on that prior. The result of the unfolding does not depend on the prior binned jet p_{T} spectrum. It does, however, depend on other prior parameters that may affect bin migration. To give some examples of such parameters:

- Particle contents of jets, their longitudinal (z) and transverse (j_{T}) fragmentation
- Contributions of the underlying event
- Distributions of jets inside each jet p_{T}, jet η bin

It is known that the jet shape slightly depends on the flavor of the initiating parton, with gluon jets being wider than quark ones. One of the selection criteria for jets – matching to the triggered Jet Patches – may bias the jet selection towards preferring to keep more of the quark-induced jets. The quark, anti-quark and gluon jets may also have a systematic difference in the total electric charge of particles inside jets. Through this, the prior Parton Distribution Functions may have some minor effect on the result of unfolding.

We are looking to evaluate the impact of the prior PDFs on the outgoing jets, without disturbing the distributions of jets inside the unfolded bins. The PDF reweighting can be performed on the existing simulation sample by applying an additional weight multiplier to

the definition of $\omega_{\text{event}}^{\text{simu.}}$ in Eq. (6.10):

$$\omega_{\text{PDF reweighting}} = \frac{f_{i_1}^{\text{target}}(x_1, Q_F^2)\, f_{i_2}^{\text{target}}(x_2, Q_F^2)}{f_{i_1}^{\text{reference}}(x_1, Q_F^2)\, f_{i_2}^{\text{reference}}(x_2, Q_F^2)} \tag{6.23}$$

A dedicated study, done as part of the current analysis, showed that, in this way, a Pythia simulation sample with the ISR, FSR and Hadronization enabled can be reweighted to match the jet p_{T} spectrum of a sample generated with a different PDF. The approach does not handle the contributions of MPI to the jet spectrum, however, that is not a critical problem for this study. The best reweighting performance was obtained by using x_1 and x_2 from `PARI(33)`, and `PARI(34)` and the factorization scale Q_F, in accordance with the hard scale choice of Perugia 2012, was taken from $\hat{p}_{\text{T}}$ =`PARI(17)`.

The result of applying the PDF reweighting procedure to construct alternative detector responses to be used in the unfolding is shown in Fig. 6.12. Although this procedure does not preserve the aforementioned distributions of jet p_{T} within bins, it shows that the uncertainty on the final cross section is small – on the order of 2% – even within this crude estimation.

6.5.5 Discrepancy in Underlying Event between data and embedding

As was discussed in Section 5.5.4, the discrepancy seen in Fig. 4.17 needs an uncertainty to be associated with it. One way to estimate it is to unfold the UE density of either the data or simulation to bring one in agreement with another. It is assumed that the discrepancy can be modeled with a $p_{\text{T}}^{\text{jet}}$-independent adjustment of the data's UE density upwards by 18% or by adjusting the embedding down by 14%. The effect of this procedure is shown in Fig. 6.13.

6.5.6 Uncertainty on the luminosity number

The relative systematic uncertainty for overall scale is assumed to be the same as was estimated for the luminosity measurement at $\sqrt{s} = 510$ GeV, which was 3.6% [68]. The "fill-by-fill" uncertainty for Vernier scans was estimated to be 3.4% (the "`run12pp200ZDCsetv1`" plot in [79]). Hence the dominant systematic uncertainty on σ_{ZDC} is 5.0%.

6.6 Comparison to theory

To achieve the goal of probing the parton distributions inside of protons, one needs to do extensive comparisons with the theory predictions. As jets convey information about the yields of outgoing partons with certain kinematics, it will need to be matched against pQCD calculations. Basic checks against the commonly accepted PDF sets should allow one to gauge the possible impact when included in the global PDF fits.

The fixed-order pQCD calculations work with jets defined at the level of individual partons. The LO pQCD matrix elements are used in most known Event Generators such as Pythia to produce events. On top of that, these codes include parton showers that effectively do a next-to-leading logarithm (NLL) resummation for soft and collinear singularities. Allowing parton

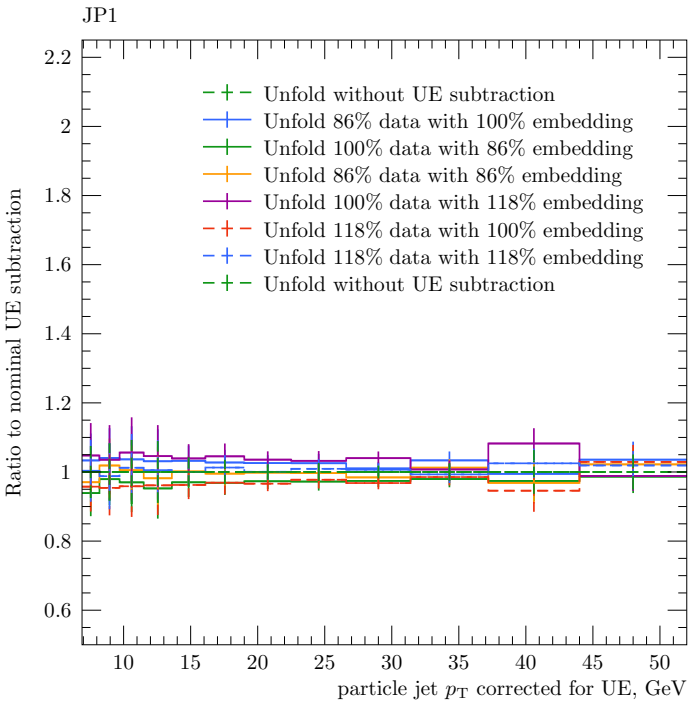
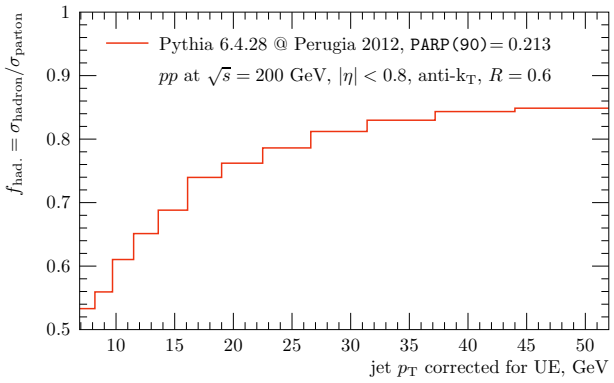
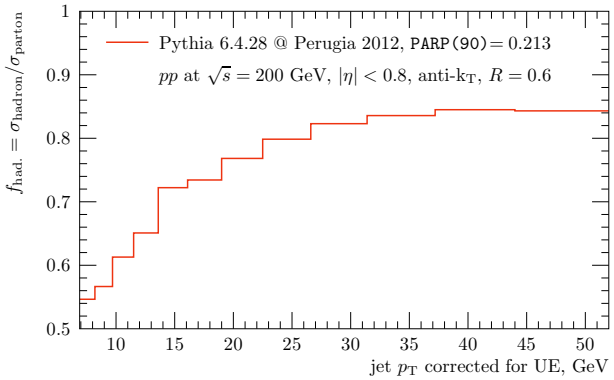


Figure 6.13: Unfolding of the data using embedding simulation samples processed with detector level UE p_T-density reduced by 14% or increased by 18% compared to the unfolding using the nominal simulation. In this case, the particle level UE density remained unchanged. Additionally, an unfolding using nominal simulation of the data processed with UE p_T-density reduced by 14% or increased by 18% is also plotted.



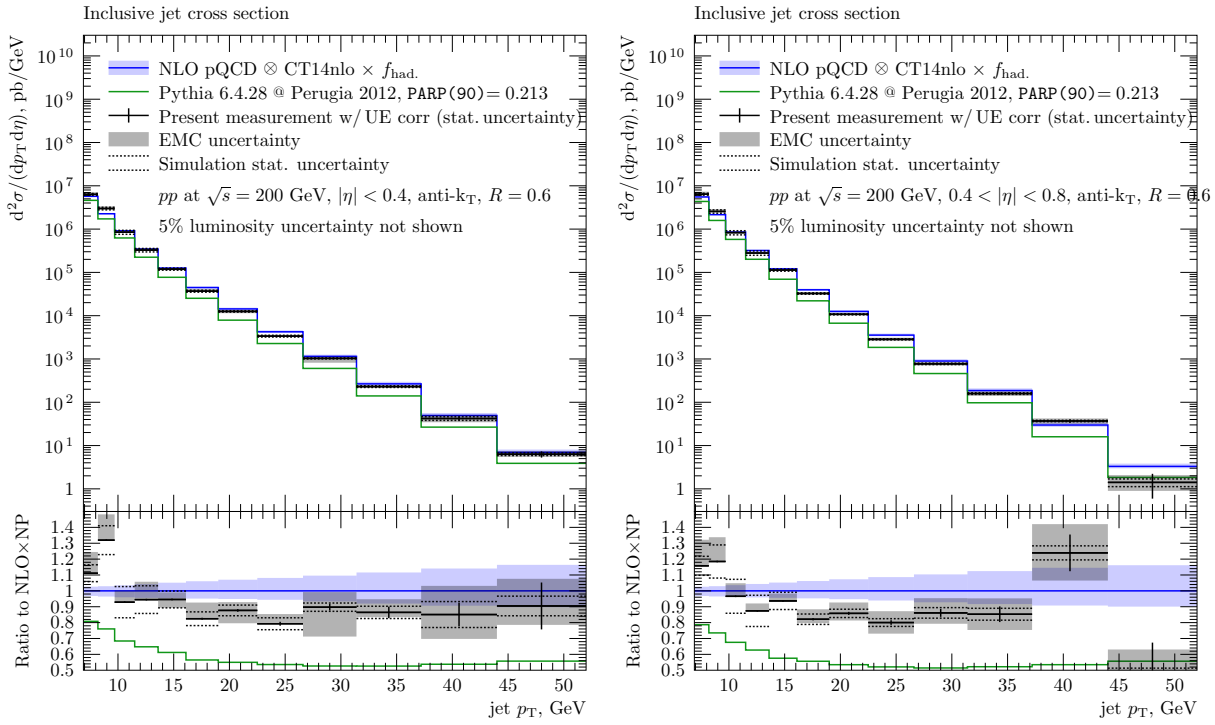
(a) $|\eta| < 0.4$



(b) $0.4 < |\eta| < 0.8$

Figure 6.14: Hadronization correction factor f_{had}.

showers to evolve the scale down to near the hadron masses is a prerequisite for simulating the hadronization, in order to get to the particle level observables. At the same time, there are NLO and even NNLO calculations that include matrix elements that are higher order in α_s. Higher orders are desirable as they allow one to reduce the theoretical uncertainties associated with the regularization and normalization scales. There is, however, a difficulty interfacing even an NLO code with NLL showers as there are overlapping contributions corresponding to real emission. Special schemes like MC@NLO [80] and POWHEG [81] can be used to avoid the double counting and allow one to "match" the NLO matrix elements to Parton Showers[4], athough such calculations are not used in the present thesis due to additional complexity of their setup.



(a) $|\eta| < 0.4$ (b) $0.4 < |\eta| < 0.8$

Figure 6.15: Differential cross section for inclusive jet production. Both the measurement and the PYTHIA simulation are have UE subtracted using the method described in Section 1.3.4. The NLO pQCD is corrected to add the effect of the hadronization using procedure described in Section 6.6.1.

6.6.1 Hadronization correction

A simple bin-by-bin procedure is used here to convert the jet cross section at the particle level to the jet cross section at the parton level. The cross section predicted by fixed order pQCD is multiplied by a ratio of the standard particle level cross section predicted by PYTHIA for our preferred tune to a parton level cross section predicted by the same event generator with the same tune, but with hadronization disabled. The parton shower is not disabled in this case, as the correction is assumed to be universal and targeted to be used with a high-order fixed order calculation. The resulting correction factors are presented in Fig. 6.14.

6.7 Results

The unfolded inclusive jet cross section, along with its uncertainties can be now presented in comparison to the PYTHIA prediction and to NLO pQCD with the hadronization correction

[4]Another tangentially related problem is the fact that a fixed-order calculation relies on possibly large and negative counter-terms to avoid numerical instabilities around integration in the infrared limit. The POWHEG scheme reshuffles the weights to make them strictly positive, thus allowing to implement the calculation into an Event Generator.

$\lvert\eta^{\text{jet}}\rvert$	$(p_{\text{T}}^{\text{jet}} - \Delta p_{\text{T}}^{\text{UE}})$, GeV	$\frac{\sigma}{\Delta p_T \, \Delta\eta} \pm \delta_{\text{stat.}} \pm \delta_{\text{simu.}} \pm \delta_{\text{JES}} \pm \delta_{\text{UE}}$, pb	$f_{\text{had.}}$
0.0 – 0.4	6.9 – 8.2	$6.34 \pm 0.02 \pm 0.30 \pm 0.76 \pm 0.46 \times 10^{6}$	0.53
	8.2 – 9.7	$2.99 \pm 0.01 \pm 0.21 \pm 0.37 \pm 0.15 \times 10^{6}$	0.56
	9.7 – 11.5	$8.47 \pm 0.05 \pm 0.90 \pm 0.67 \pm 0.43 \times 10^{5}$	0.61
	11.5 – 13.6	$3.27 \pm 0.02 \pm 0.30 \pm 0.39 \pm 0.16 \times 10^{5}$	0.65
	13.6 – 16.1	$1.19 \pm 0.01 \pm 0.07 \pm 0.07 \pm 0.05 \times 10^{5}$	0.69
	16.1 – 19.0	$3.68 \pm 0.03 \pm 0.19 \pm 0.46 \pm 0.11 \times 10^{4}$	0.74
	19.0 – 22.5	$1.26 \pm 0.01 \pm 0.05 \pm 0.12 \pm 0.05 \times 10^{4}$	0.76
	22.5 – 26.6	$3.36 \pm 0.05 \pm 0.16 \pm 0.26 \pm 0.09 \times 10^{3}$	0.79
	26.6 – 31.4	$1.03 \pm 0.02 \pm 0.03 \pm 0.21 \pm 0.04 \times 10^{3}$	0.81
	31.4 – 37.2	$2.30 \pm 0.10 \pm 0.10 \pm 0.22 \pm 0.06 \times 10^{2}$	0.83
	37.2 – 44.0	$4.22 \pm 0.37 \pm 0.41 \pm 0.90 \pm 0.17 \times 10^{1}$	0.84
	44.0 – 52.0	$6.30 \pm 1.03 \pm 0.43 \pm 1.19 \pm 0.18 \times 10^{0}$	0.85
0.4 – 0.8	6.9 – 8.2	$6.40 \pm 0.02 \pm 0.33 \pm 0.91 \pm 0.43 \times 10^{6}$	0.55
	8.2 – 9.7	$2.55 \pm 0.01 \pm 0.22 \pm 0.33 \pm 0.16 \times 10^{6}$	0.57
	9.7 – 11.5	$8.26 \pm 0.04 \pm 0.91 \pm 0.71 \pm 0.53 \times 10^{5}$	0.61
	11.5 – 13.6	$2.80 \pm 0.02 \pm 0.32 \pm 0.15 \pm 0.16 \times 10^{5}$	0.65
	13.6 – 16.1	$1.13 \pm 0.01 \pm 0.07 \pm 0.06 \pm 0.05 \times 10^{5}$	0.72
	16.1 – 19.0	$3.25 \pm 0.03 \pm 0.13 \pm 0.25 \pm 0.13 \times 10^{4}$	0.73
	19.0 – 22.5	$1.08 \pm 0.01 \pm 0.03 \pm 0.09 \pm 0.03 \times 10^{4}$	0.77
	22.5 – 26.6	$2.85 \pm 0.05 \pm 0.09 \pm 0.25 \pm 0.10 \times 10^{3}$	0.80
	26.6 – 31.4	$7.72 \pm 0.22 \pm 0.30 \pm 0.83 \pm 0.23 \times 10^{2}$	0.82
	31.4 – 37.2	$1.59 \pm 0.09 \pm 0.07 \pm 0.19 \pm 0.03 \times 10^{2}$	0.84
	37.2 – 44.0	$3.70 \pm 0.34 \pm 0.13 \pm 0.54 \pm 0.12 \times 10^{1}$	0.84
	44.0 – 52.0	$1.41 \pm 0.82 \pm 0.28 \pm 0.67 \pm 0.06 \times 10^{0}$	0.84

Table 6.6: Numerical values for the differential Inclusive Jet cross section for proton-proton collisions at $\sqrt{s} = 200$ GeV corresponding to Fig. 6.15. The last column contains numerical values of the hadronization correction factor values from Fig. 6.14. Those are suggested for use in comparisons of the present measurement to the fixed-order pQCD calculations.

applied. The result is presented graphically in Fig. 6.15 and as numbers in Table 6.6.

In the next chapter an interpretation of the present result will be discussed.

Chapter 7

QCD analysis

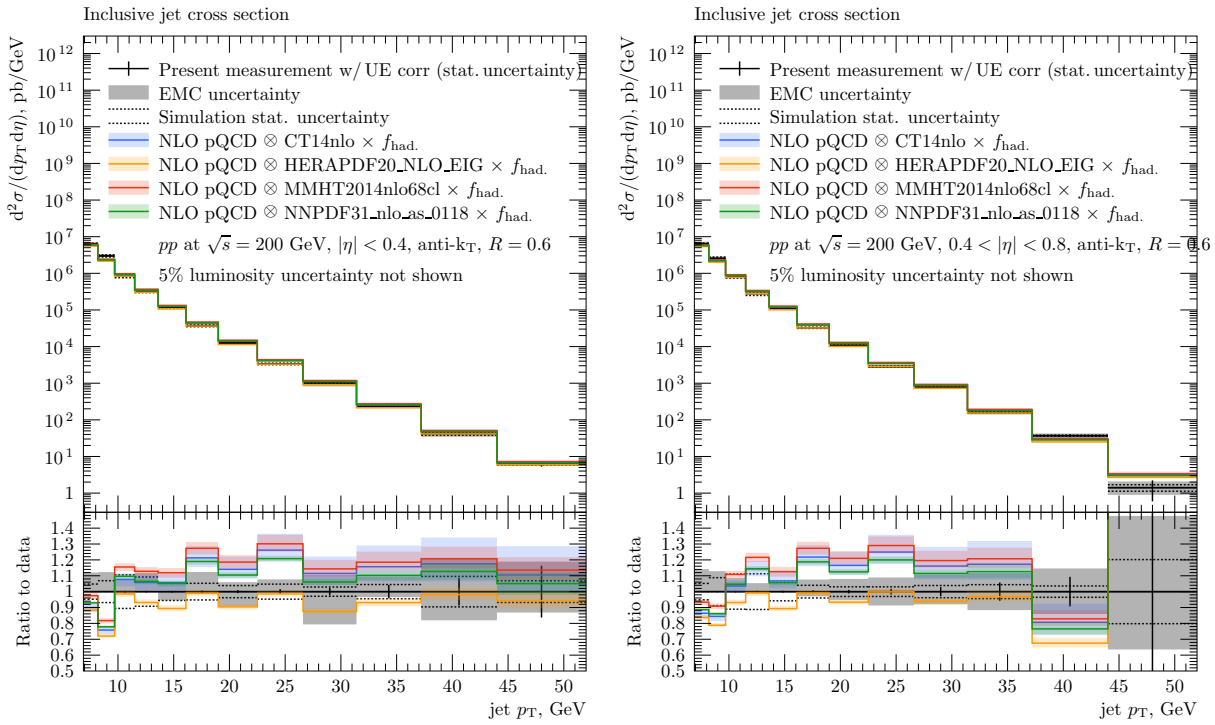
QCD provides the collinear factorization framework that connects observable cross sections to the Parton Distribution Functions. The present measurement alone is not sufficient to reconstruct the PDFs for several reasons. A typical PDF fit requires at least a Deep Inelastic Scattering (DIS) data set, for which in Neutral Current processes (NC; $e+p \to \gamma/Z \to e+X$) one can pinpoint linear combinations of quark and anti-quark distributions with good kinematic resolution in x and Q^2 (limited by radiative corrections). The DIS data set on Charged Current processes (CC; $l+p \to W \to \nu_l + X$ or $\nu_l + p \to \gamma/Z \to l + X$) allows one to further disentangle up-type and down-type quark/antiquark distributions. Hadron collider data on $W^{\pm}$ boson production cross sections have sensitivity between the quark generations allowing one to constrain the strange quark distribution. Measurements of the jet production cross sections in ep and pp collisions have a direct sensitivity to the gluon PDF.

A common method for benchmarking a new dataset in a global fit is to reweight one of the existing PDF fits using the information on its correlated errors. The specific procedure depends on the details of the definition used to calculate uncertainties in a given PDF fit: for the fits using the Hessian approach (as used by CTEQ, xFitter, MMHT collaborations) the formalism presented in [23] can be applied, while fits that use Monte Carlo sampling (mainly used by the NNPDF Collaboration) can be reweighted using a Bayesian formalism such as [82]. In this thesis, the impact on the fit is assessed using the xFitter [83] framework to conduct a full fit using the present inclusive jet cross section measurements in addition to the combined data [84] on cross sections for the DIS processes measured by the H1 and Zeus experiments at the HERA ep collider. This should allow one to get most precise estimates for the updated PDF values, and more importantly, values for their uncertainties.

7.1 General comparison to NLO pQCD

Impact on the PDF fits can be estimated by comparing predictions of pQCD for different existing PDF fits. The CT14 [85], HERAPDF2.0 [84], MMHT2014 [86] and NNPDF3.1 [87] PDF sets were used for comparisons shown in Fig. 7.1.

The data favor the predictions for the HERAPDF2.0 set that has an x-dependency trend that is different from the trend for the other chosen PDF sets. The latter observation can be explained by looking at the gluon component of the PDF sets presented in Fig. 7.2, which shows



(a) $|\eta| < 0.4$ (b) $0.4 < |\eta| < 0.8$

Figure 7.1: NLO pQCD predictions for several modern PDF sets compared to the present measurement. Error bands indicate uncertainty propagated from the uncertainty of the PDF sets.

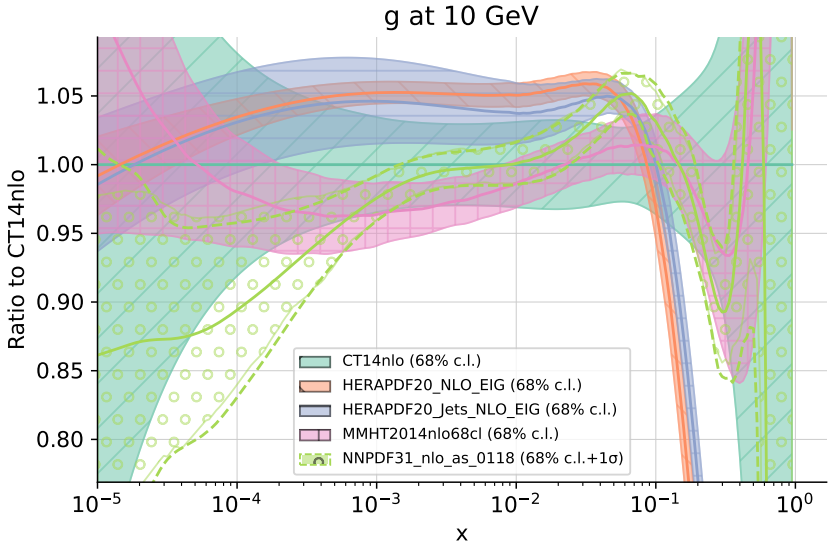


Figure 7.2: Comparison of the gluon PDF as a function of x for several modern PDF sets at a scale $Q = 10$ GeV.

a drastically reduced gluon contribution at high x[1]. Interestingly, the HERAPDF2.0Jets [84] PDF set, that is supposed to have a direct sensitivity to the gluon PDF by the virtue of including data on jet production at HERA, has a similar trend as the HERAPDF2.0.

The poor agreement of the present data with NLO pQCD for PDFs other than HERAPDF2.0 further motivates use of the xFitter framework or otherwise reweighting that particular PDF set.

7.2 Procedure for PDF fitting

A general outline of the procedure for determining a PDF from a fit to the data goes as follows:

- Choose a functional form for all PDFs at an initial scale (the standard value from HERAPDF2.0, $Q^2 = 1.9$ GeV^2 is used here) dependent on a set of numeric fit parameters a (to be described in Section 7.2.1)
- Choose initial values of the PDF parameters
- Numerically evolve the PDFs using DGLAP evolution [88–91] to the Q^2 scale(s) needed by the theory to predict the experimental data
- Calculate theoretical predictions m from the values at the current evolved PDF (to be described in Section 7.2.2)
- Compare predictions m for the evolved PDF to the data μ, calculate χ^2 as described in Section 7.2.3
- Iterate steps 3, 4 and 5 for different values of numeric parameters to determine derivatives $\partial\chi^2/\partial a_i$ of χ^2 with respect to the parameters a as needed by the `MIGRAD` minimization algorithm from `MINUIT` [92]
- Allow `MIGRAD` to iterate steps 3, 4, 5 and 6 and update values of a until a minimum of χ^2 is reached

Once the fit is done, the values of the parameters $a = a_{\text{best-fit}}$ corresponding to the minimum of the χ^2 will determine the resulting PDF set. The Hessian matrix H is calculated using a method from [93]. Eigenvectors of the Hessian matrix δa define a correlated uncertainty on the vector a using a tolerance $T \equiv 1 = \Delta\chi^2 = \sum_{ij} \delta a_i H_{ij} \delta a_j$. The PDF sets corresponding to updating the parameters from $a_{\text{best-fit}}$ to $a_{\text{best-fit}} \pm \delta a$ are called "Hessian error vectors", and they are used to describe the uncertainties of a PDF set.

[1] A check to see if the gluon distribution is responsible for the shape modification can be done by calculating the prediction of NLO pQCD for a PDF set that uses the quark distributions from HERAPDF and the gluon distribution from non-HERAPDF. While such a PDF set violates the momentum sum rule, this is still a valid numerical check. This result shows that the gluon distribution is indeed the primary cause for the effect.

7.2.1 Fit parametrization

The standard parametrization of HERAPDF2.0 [84] is the following:

$$x\,g(x) = A_g\,x^{B_g}(1-x)^{C_g} - A'_g\,x^{B'_g}(1-x)^{C'_g} \tag{7.1}$$
$$x\,u_v(x) = A_{u_v}\,x^{B_{u_v}}(1-x)^{C_{u_v}}\,(1+E_{u_v}x^2) \tag{7.2}$$
$$x\,d_v(x) = A_{d_v}\,x^{B_{d_v}}(1-x)^{C_{d_v}} \tag{7.3}$$
$$x\,\bar{U}(x) = A_{\bar{U}}\,x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}\,(1+D_{\bar{U}}x) \tag{7.4}$$
$$x\,\bar{D}(x) = A_{\bar{D}}\,x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}} \tag{7.5}$$
$$\tag{7.6}$$

Then the normalization parameters A_g, A_{u_v}, A_{d_v} are fixed using the momentum sum rule and valence quark sum rules:

$$1 = \int \mathrm{d}x\; x\left(\sum_i (q_i(x) + \bar{q}_i(x)) + g(x)\right) \tag{7.7}$$
$$2 = \int \mathrm{d}x\; u_v(x) \tag{7.8}$$
$$1 = \int \mathrm{d}x\; d_v(x) \tag{7.9}$$
$$\tag{7.10}$$

Other constraints are set on the parameters: $B_{\bar{U}} = B_{\bar{D}}$, $C'_g = 25$, $A_{\bar{U}} = A_{\bar{D}}(1-f_s)$ with $f_s = 0.4$. This leaves 14 free parameters in the fit.

7.2.2 Theory

In order to calculate the theoretical predictions m^i for a given set of PDFs at specified order, one needs a pQCD calculation matching to the same order. In this analysis NLO pQCD theory implemented in the NLOJet++ code [94] is used to obtain the parton level jet cross sections. This code is using a variant of the Catani-Seymour subtraction [95] that relies on numerical integration to integrate over the phase space. A PDF fit generally requires one to evaluate the theory prediction for many varying PDF inputs, which is impractical to do by repeating the slow integration each time. One way to improve the speed of the calculation is based on using the QCD factorization theorem [96]. In this method one represents the cross section as a sum of PDFs interpolated on a grid of discrete x-Q values and coupling $\alpha_s(Q)$ interpolated on a grid of just the Q values with cross section kernels. The kernels are specific to a particular process, energy and jet observable, but can be calculated once (100 CPU days) and be quickly convoluted with an arbitrary PDF set ($\ll 1$ CPU second) to get a result for the cross section. The FastNLO tables are generated for this measurement and are available at [97].

The partonic level jet cross section from the interpolated fixed order pQCD is further corrected to particle level using the multiplicative correction described in Section 6.6.1. The correction is assumed to be weakly dependent on the PDF set used.

Parameter	HERA only	HERA+pp200
'Adbar'	0.1054 ± 0.0032	0.1089 ± 0.0058
'Adv'	1.0000	1.0000
'Ag'	1.0000	1.0000
'Agp'	0.237 ± 0.021	0.141 ± 0.039
'Auv'	1.0000	1.0000
'Bdbar'	−0.1723 ± 0.0036	−0.1680 ± 0.0062
'Bdv'	0.805 ± 0.037	0.827 ± 0.091
'Bg'	−0.0135 ± 0.0018	−0.0257 ± 0.0055
'Bgp'	−0.1666 ± 0.0095	−0.239 ± 0.046
'Buv'	0.7138 ± 0.0080	0.700 ± 0.031
'Cdbar'	4.86 ± 0.54	8.5 ± 2.1
'Cdv'	4.08 ± 0.16	4.08 ± 0.41
'Cg'	9.09 ± 0.26	7.00 ± 0.27
'Cgp'	25.00	25.00
'Cubar'	8.06 ± 0.13	7.04 ± 0.78
'Cuv'	4.841 ± 0.040	4.807 ± 0.087
'DbarToS'	1.0000	1.0000
'Dubar'	11.9 ± 1.0	8.4 ± 2.3
'Euv'	13.40 ± 0.57	13.4 ± 2.0
'fs'	0.4000	0.4000
Fit status	not-a-fit	not-a-fit

Table 7.1: PDF parameters resulting from PDF fits to the HERA data only and to the HERA data with addition of the present measurement.

7.2.3 Definition of χ^2 in the fit

Equation 32 from [84] reads:

$$\chi^2_{\text{exp}} = \sum_i \frac{[m^i - \sum_j \gamma^i_j m^i s_j - \mu^i]^2}{\delta^2_{i,\text{stat}}\mu^i m^i + \delta^2_{i,\text{uncor}}(m^i)^2} + \sum_j s_j^2 + \sum_i \log \frac{\delta^2_{i,\text{stat}}\mu^i m^i + \delta^2_{i,\text{uncor}}(m^i)^2}{(\delta^2_{i,\text{stat}} + \delta^2_{i,\text{uncor}})(\mu^i)^2} \quad (7.11)$$

The first term includes a deviation in units of "sigma" of the theoretical prediction m^i from the measured data point μ^i. It is written under the assumption that the actual uncertainty has to be calculated from the true value μ^i, hence, the statistical uncertainty scales just as the Poisson variable $\mu^i \sim \sqrt{N}$ and the uncorrelated uncertainty scales linearly, assuming a fixed relative uncertainty $\delta_{i,\text{uncor}}$. The s_j are so-called "nuisance parameters", each showing by how many "sigmas" (towards the penalty accumulated in the second term) the relative correlated uncertainty vector γ^i_j is shifted. The third term with the logarithm originally introduced in [98] serves the purpose of minimising the bias associated with the presence of the variable quantity, such as m^i in the denominator of the first term (acting as the Gaussian width).

Dataset	HERA only	HERA+pp200
HERA1+2 NCep 820	68 / 70	69 / 70
HERA1+2 NCep 460	217 / 204	220 / 204
HERA1+2 CCep	43 / 39	46 / 39
HERA1+2 NCem	222 / 159	224 / 159
HERA1+2 CCem	54 / 42	53 / 42
HERA1+2 NCep 575	219 / 254	222 / 254
HERA1+2 NCep 920	439 / 377	454 / 377
STAR pp200 jets	-	66 / 24
Correlated χ^2	86	84
Log penalty χ^2	+8.4	-6.81
Total χ^2 / dof	1357 / 1131	1431 / 1155
χ^2 p-value	0.00	0.00

Table 7.2: $\chi^2/N_{\text{d.o.f.}}$ for PDF fits to the HERA data only and to the HERA data with addition of the present measurement.

7.2.4 Results

A fit to the HERA data only using xFitter 2.2.0 agrees reasonably well with the published HERAPDF2.0 PDF and will serve as a baseline. Adding the present measurement to the fit requires three components: theory tables (described in Section 7.2.2), measured values (given in Table 6.6) and their uncertainties. The statistical uncertainties of the data and the simulation (described in Sections 6.4 and 6.5.1) were treated as fully uncorrelated between bins, whereas the leading systematic uncertainties for EMC response (Section 6.5.2) and luminosity (Section 6.5.6) were treated as fully correlated. The other uncertainties are neglected in this fit.

The original HERAPDF2.0 fit at NLO has a $\chi^2/N_{\text{d.o.f.}} = 1357/1131$ [84], and, as seen in Table 7.2, the present data add an additional 24 degrees of freedom and increase χ^2 by 66 units.

A notable change in the PDF fit after inclusion of the present data happens to the gluon PDF, as is shown in Fig. 7.3. The change to the inclusive jet cross section predicted by NLO pQCD is shown in Fig. 7.4.

When one is interested in quantitatively comparing the ability of different datasets to constrain PDFs, a direct comparison of the relative uncertainties like $\Delta g(x, Q^2)/g(x, Q^2)$ is subject to variations due to different data resulting in fits with different values of the denominator $g(x, Q^2)$. One way to overcome that is to replace data with pseudo-data by replacing the point values with their theory predictions for the reference PDF. This technique is used to compare the impact of the present inclusive jet cross section measurement at RHIC to the impact of the inclusive jet cross section measurements from the Tevatron [99, 100]. The result is given in Fig. 7.5 and shows that the present measurement performs better at high x than those past measurements at the Tevatron.

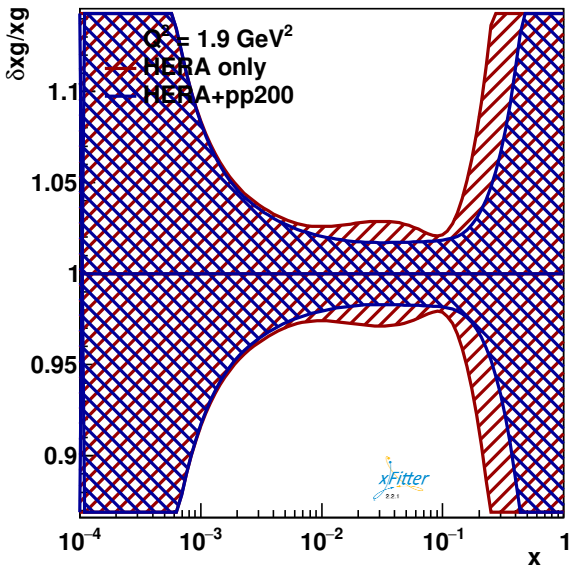
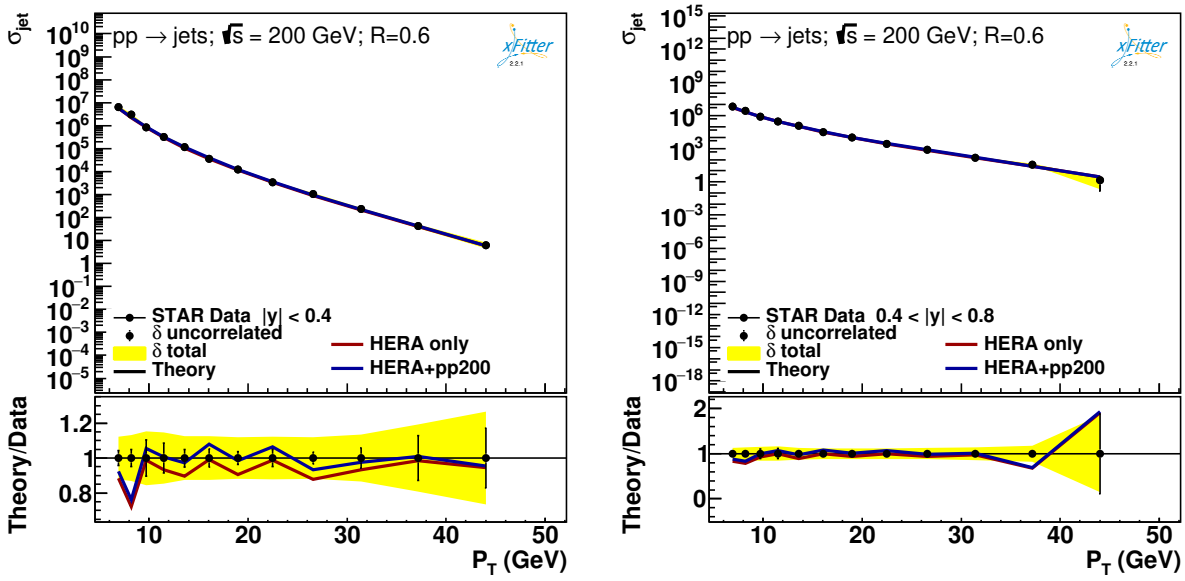


Figure 7.3: Relative uncertainties on the values of $g(x, Q^2)$ as a function of x at the parametrization scale $Q^2 = 1.9$ GeV^2 for PDF fits to the HERA data only and to the HERA data with addition of the present measurement.



(a) $|\eta| < 0.4$ (b) $0.4 < |\eta| < 0.8$

Figure 7.4: Inclusive jet cross section predicted by NLO pQCD for the PDF sets resulting from the fits with and without the present data (referred to as "pp200").

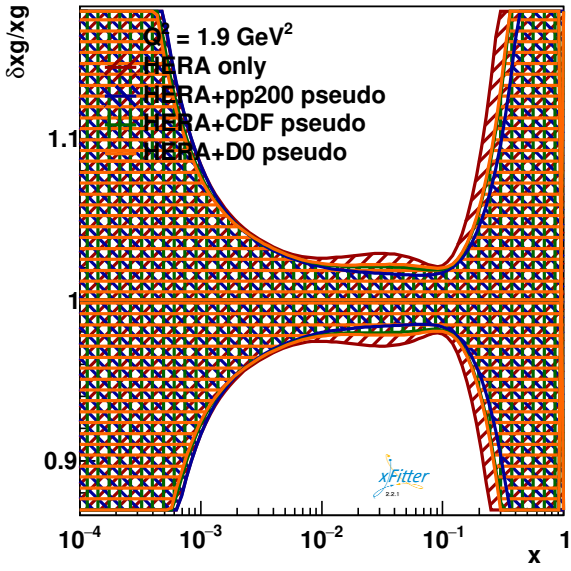


Figure 7.5: Relative uncertainties on the values of $g(x, Q^2)$ as a function of x at the parametrization scale $Q^2 = 1.9$ GeV^2 for pseudo-data fits using the HERA data alone (HERAPDF2.0-like) and in combination with pseudo-data derived from measurements at STAR, CDF and D0.

Chapter 8

Conclusions

The cross section for inclusive jet production in unpolarized proton-proton collisions has been measured in a unique kinematic region, at $\sqrt{s} = 200$ GeV, for the first time, with the focus on precision and careful treatment of the systematic uncertainties. This measurement is complementary for use in studies of the initial proton state and the final state interactions in jet physics to the previous measurements at the Tevatron and LHC.

The goal for this measurement is to provide an improved constraint on the gluon PDF at large values of x. This has been demonstrated in comparison to previous inclusive jet cross section measurements done at the Tevatron that are included in most modern global fits. The presented measurement uses the anti-k_{T} algorithm which, unlike the mid-point cone algorithm used at Tevatron, is collinear and infrared safe to all orders in perturbation theory. Because of that, the present measurement will continue to be useful as we advance into the era of NNLO and N^{3}LO theoretical calculations.

The present analysis suggests that this measurement has the potential to improve the gluon PDF uncertainties in the global fits, contingent on knowing the value of the hadronization correction. While the theoretical uncertainty on this correction was estimated in this work, its size is subjective. Should subsequent theoretical studies reveal the uncertainty to be much larger, the present measurement can be used as a basis for tuning MC simulations by leveraging the existing precision of the PDF sets. Improvement in understanding of the hadronization effects for jets at RHIC energies would be valuable for interpreting jet measurements at the future Electron Ion Collider.

Appendix A

Collider luminosity

The instantaneous luminosity $\mathcal{L}$ of a scattering system $\mathcal{B}$ (specified in terms of experimental conditions, e.g., density and flux of incoming beams) quantifies for every interaction process $\mathcal{P}$ (specified in terms of the final state conditions, e.g., the momenta and the species of particles scattered or produced in the interaction) a ratio between its cross section $\sigma^{\mathcal{P}}$ and interaction rate $\frac{dN^{\mathcal{B};\mathcal{P}}}{\mathrm{d}t}$:

$$\mathcal{L}^{\mathcal{B}} \equiv \frac{1}{\sigma^{\mathcal{P}}} \frac{dN^{\mathcal{B};\mathcal{P}}}{\mathrm{d}t}, \tag{A.1}$$

The Si unit for luminosity is therefore $\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Using the quantities as measured in the rest frame of one of the colliding beams of density $\rho^{\mathcal{B}\text{-target}}$, the cross section $\sigma^{\mathcal{P}}$ can be defined in terms of incident beam's flux $\Phi^{\mathcal{B}\text{-incident}}$, scattering rate $\frac{dN^{\mathcal{B};\mathcal{P}}}{\mathrm{d}t}$ and interaction region volume V as:

$$\sigma^{\mathcal{P}}\ \rho^{\mathcal{B}\text{-target}}\ \Phi^{\mathcal{B}\text{-incident}}\ V = \frac{dN^{\mathcal{B};\mathcal{P}}}{\mathrm{d}t} \tag{A.2}$$

A cross section quantifies the probability for interaction, however it has a geometrical interpretation. One example is the classical scattering of a point-like object on a solid black body for which the geometrical cross section in the plane transverse to the incidence axis matches the interaction cross section. Another example is the scattering at various possible impact parameters (i.e., transverse distances with respect to the scattering centers) for microscopic systems. Hence a cross section is given in units of area.

Combining Eq. (A.2) with Eq. (A.1) and integrating over time and allowing for non-uniform densities in the interaction volume one gets a general expression for the integrated luminosity:

$$\mathcal{L}_{\text{int.}} \equiv \int \mathcal{L}\ \mathrm{d}t = \int \rho_1 \rho_2 v_{12}\ \mathrm{d}V\mathrm{d}t \tag{A.3}$$

Both $\mathcal{L}_{\text{int.}}$ and $\mathrm{d}V\mathrm{d}t$ are Lorentz-invariants. Substituting a Lorentz-invariant expression for the luminosity density $\rho_1\rho_2 v_{12}$ one gets [101]:

$$\mathcal{L}_{\text{int.}} = \int \rho_1 \rho_2 \sqrt{(\vec{v}_1 - \vec{v}_2)^2 - \frac{(\vec{v}_1 \times \vec{v}_2)^2}{c^2}}\ \mathrm{d}V\mathrm{d}t \tag{A.4}$$

The angle between $\vec{v}_1$ and $-\vec{v}_2$ is known as the crossing angle. If its value is negligibly small, the latter square root evaluates for relativistic head-on collisions to simply $2c$.

A.1 Vernier scan

The measurement of the luminosity for this analysis was performed using a "Van der Meer scan" [102] (also known as a Vernier scan). At RHIC this was done by changing beam trajectories to achieve a displacement of the colliding "Blue" and "Yellow" beams with respect to each other in the horizontal and vertical directions.

A.1.1 Beam position

The x and y coordinates of the beams are measured using the Beam Position Monitors (BPMs) [103] located near DX magnets[1] located at both sides of the STAR's Intersection Point (IP). The location of the beams can be then interpolated as a mean of the measurements at both ends:

$$\begin{aligned} x_{i;\mathrm{IP}} &= \frac{1}{2}(x_{i;\mathrm{BPM\ in}} + x_{i;\mathrm{BPM\ out}}) \\ y_{i;\mathrm{IP}} &= \frac{1}{2}(y_{i;\mathrm{BPM\ in}} + y_{i;\mathrm{BPM\ out}}) \end{aligned} \tag{A.5}$$

This way one can monitor beam positions for each of the two beams (indexed with the index i) and cross-check the displacements applied during a Vernier scan.

A.1.2 Luminosity monitoring

Another component needed to make the Vernier scan work is an ability to track the relative changes in the luminosity. In an experiment this is typically implemented by setting up a trigger that would detect an occurrence of a certain physical interaction process $\mathcal{P}_{\mathrm{mon.}}$ corresponding to a fixed interaction cross section $\sigma^{\mathcal{P}_{\mathrm{mon.}}}$. Then the firing rate $\frac{\mathrm{d}N^{\mathcal{P}_{\mathrm{mon.}}}}{\mathrm{d}t}$ of that trigger would be proportional to the instantaneous luminosity $\mathcal{L}$.

For a luminosity monitor trigger to serve its purpose well, it should satisfy some experimental requirements: the background has to be small or predictable, the deadtime has to be minimal, and the energy or momentum thresholds and the acceptance should be stable. For this analysis the luminosity monitoring was done using the rate of coincidence of hits in the East and West ZDC detectors. Having ZDC detectors on each side of STAR makes the requirement of coincidence suppress some background. It must also have a uniform acceptance and response across the observed interaction diamond. Situated right along the beam axis the ZDC detectors naturally satisfy the latter requirement.

There are, however, some non-linearities present for a coincidence-based trigger at high luminosity. For one, such a trigger can not distinguish if single or multiple interactions have

[1] A set of "DX" and "D0" dipole magnets is located at ≈ 10 m and ≈ 23 m away from every Intersection Point of the RHIC [104]. The purpose of the first D0 magnets is to steer beams from the beamline into the first DX magnets, where they are focused towards a collinear path onto the incident beams. The second DX magnet steers the beam back to the D0 magnet which directs it back into the beam line.

occurred in a given bunch crossing. Additionally, when we observe hits at both ends at the same time, which would usually indicate an interaction of interest – a true coincidence, but may be also due to a coincidence of backgrounds firing both detectors at the same time. These two effects are small for proton-proton collisions at RHIC at $\sqrt{s} = 200$ GeV that are low-luminosity in comparison to proton-proton collisions at $\sqrt{s} = 510$ GeV. They were addressed via a correction for "multiples" and "accidentals" that is described in [105].

There are at least three [106] systems in STAR that allow tracking and recording the ZDC coincidence rate [107]. The RICH scaler boards are hardware counters that can accumulate signal counts and report them at regular intervals. Unfortunately, the raw RICH scaler data with per second resolution is not available for the dataset for pp collisions at $\sqrt{s} = 200$ GeV in year 2012, although the rates are also stored in the experiment conditions database, but only at 15 second intervals (i.e., with 14 second gaps). Data from the RHIC scaler system is available at 1 second intervals. However, it uses different digitization hardware for the analog signals from the ZDC and thus provides a rate that is different from the RICH scaler. Thus choice of a scaler affects the definition of what is considered to be the ZDC cross section. The third set of scaler logic is present in the TCIM unit and monitors the ZDC trigger fire condition rates, which relies upon both ADC and TAC timing information, as would be used for the ZDCMB trigger.

A.1.3 Luminosity in Vernier scans

Rewriting Eq. (A.4) for collinear beams colliding with horizontal and vertical displacements Δx and Δy, we get an expression for luminosity given as the following overlap integral:

$$\mathcal{L}_{\text{int.}} = 2c \int \rho_1(x, y, z, t)\, \rho_2(x - \Delta x, y - \Delta y, z, -t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}t \tag{A.6}$$

for which the densities along x, y and z of a single bunch are often assumed to be Gaussian:

$$\rho_i(x, y, z, t) = \frac{N_i}{\sqrt{(2\pi)^3}\, \sigma_{x;i}\, \sigma_{y;i}\, \sigma_{z;i}} \exp\left(-\frac{x^2}{2\sigma_{x;i}^2} - \frac{y^2}{2\sigma_{y;i}^2} - \frac{(z - ct)^2}{2\sigma_{z;i}^2} \right) \tag{A.7}$$

and so the integration in Eq. (A.6) along the x and y axes is performed as:

$$\begin{aligned} \frac{1}{2\pi\sigma_{x;1}\sigma_{x;2}} \int \exp\left(-\frac{x_1^2}{2\sigma_{x;1}^2} - \frac{(x_1 + \Delta x)^2}{2\sigma_{x;2}^2} \right) \mathrm{d}x_1 = \\ = \frac{1}{\sqrt{2\pi(\sigma_{x;1}^2 + \sigma_{x;2}^2)}} \exp\left(-\frac{(\Delta x)^2}{2(\sigma_{x;1}^2 + \sigma_{x;2}^2)} \right) \end{aligned} \tag{A.8}$$

and integration over time t of a single bunch crossing gives:

$$\begin{aligned} \frac{1}{2\pi\sigma_{z;1}\sigma_{z;2}} \int \exp\left(-\frac{(z - ct)^2}{2\sigma_{z;1}^2} - \frac{(z + ct)^2}{2\sigma_{z;2}^2} \right) \mathrm{d}z\, \mathrm{d}t = \\ = \frac{1}{c\sqrt{2\pi(\sigma_{z;1}^2 + \sigma_{z;2}^2)}} \int \exp\left(-\frac{4z^2}{2(\sigma_{z;1}^2 + \sigma_{z;2}^2)} \right) \mathrm{d}z \end{aligned} \tag{A.9}$$

Averaging over many beam revolutions involves summing over all the colliding bunch pairs in the beams, and one arrives at:

$$\left\langle \frac{\mathrm{d}\mathcal{L}(\Delta x, \Delta y)}{\mathrm{d}t} \right\rangle = 2\frac{f_{\mathrm{rev}} \sum_i N_{1;i} N_{2;i}}{\sqrt{(2\pi)^3}\, \Sigma_x\, \Sigma_y\, \Sigma_z} \exp\left(-\frac{\Delta x^2}{2\Sigma_x^2} - \frac{\Delta y^2}{2\Sigma_y^2} \right) \int \exp\left(-\frac{4z^2}{2\Sigma_z^2} \right) \mathrm{d}z = \\ = \frac{f_{\mathrm{rev}} \sum_i N_{1;i} N_{2;i}}{2\pi\, \Sigma_x\, \Sigma_y} \exp\left(-\frac{\Delta x^2}{2\Sigma_x^2} - \frac{\Delta y^2}{2\Sigma_y^2} \right) \quad \text{(A.10)}$$

where $N_{1;i}$ and $N_{2;i}$ are the total numbers of particles in the pairs of colliding bunches in the two beams, and where shorthand for the terms $\Sigma_x \equiv (\sigma_{x;1}^2 + \sigma_{x;2}^2)$, $\Sigma_y \equiv (\sigma_{y;1}^2 + \sigma_{y;2}^2)$ and $\Sigma_z \equiv (\sigma_{z;1}^2 + \sigma_{z;2}^2)$ are used.

The quantities of interest are the maximum possible instantaneous luminosity during the Vernier scan

$$\left\langle \frac{\mathrm{d}\mathcal{L}(0,0)}{\mathrm{d}t} \right\rangle = \frac{f_{\mathrm{rev}} \sum_i N_{1;i} N_{2;i}}{2\pi\, \Sigma_x\, \Sigma_y} \quad \text{(A.11)}$$

and instantaneous luminosity integrated over displacement along an axis of interest (x-axis used as a specific example):

$$\int \left\langle \frac{\mathrm{d}\mathcal{L}(\Delta x, 0)}{\mathrm{d}t} \right\rangle \mathrm{d}(\Delta x) = \frac{f_{\mathrm{rev}} \sum_i N_{1;i} N_{2;i}}{\sqrt{2\pi}\, \Sigma_y}. \quad \text{(A.12)}$$

Hence, by measuring the luminosity monitor trigger rate $\frac{\mathrm{d}N^{\mathcal{P}_m}}{\mathrm{d}t}(\Delta x, \Delta y) \sim \mathcal{L}(\Delta x, \Delta y)$ for several known beam displacements, one can determine the combined beam widths Σ_x and Σ_y:

$$\Sigma_x = \frac{\int \left\langle \frac{\mathrm{d}\mathcal{L}(\Delta x, 0)}{\mathrm{d}t} \right\rangle \mathrm{d}(\Delta x)}{\sqrt{2\pi} \left\langle \frac{\mathrm{d}\mathcal{L}(0,0)}{\mathrm{d}t} \right\rangle} \quad \text{(A.13)}$$

$$\Sigma_y = \frac{\int \left\langle \frac{\mathrm{d}\mathcal{L}(0, \Delta y)}{\mathrm{d}t} \right\rangle \mathrm{d}(\Delta y)}{\sqrt{2\pi} \left\langle \frac{\mathrm{d}\mathcal{L}(0,0)}{\mathrm{d}t} \right\rangle} \quad \text{(A.14)}$$

The resulting values for Σ_x and Σ_y can be then substituted back into Eq. (A.11) or Eq. (A.10).

The stability of the RHIC beam conditions is not guaranteed even within a single RHIC fill. Hence, the luminosity determined for the beam conditions that were in place during the Vernier scan does not by itself determine the luminosity delivered during data-taking that is used for the physics cross section measurement.

Hourglass correction

The formalism of beam dynamics describes the transverse beam size using the following formula:

$$\sigma_i(s) = \sqrt{\epsilon_i\, \beta_i(s)} \quad \text{(A.15)}$$

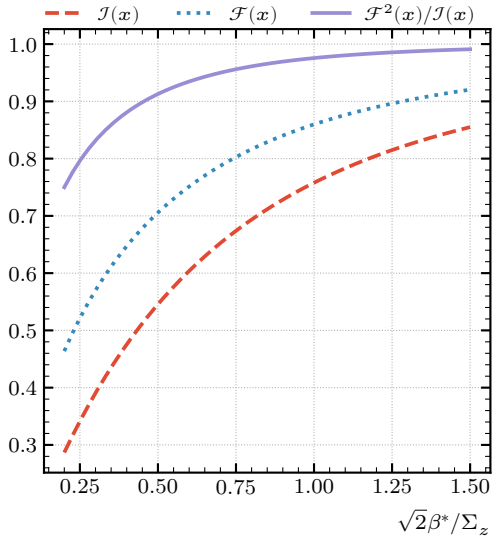


Figure A.1: Form factors $\mathcal{J}$, $\mathcal{F}$ and $\mathcal{F}^2/\mathcal{J}$ as a function of $\sqrt{2}\beta^*/\Sigma_z \approx \beta^*/\sigma_z$.

where ϵ_i is the beam emittance along axis i, and $\beta_i(s)$ is a "beta function" parametrized along the nominal beam trajectory. An important result of beam dynamics is that for a beam a "drift space" (without presence of any magneto-optics) behaves according to:

$$\beta_i(z) = \beta_i^* + \frac{z^2}{\beta_i^*} \tag{A.16}$$

where β_i^* is the value[2] of the beta function at a focal point $z = 0$:

$$\beta_i^* = \beta_i(0) \tag{A.17}$$

This effect in turn modifies the overlap density to look like

$$\begin{aligned}\left\langle \frac{\mathrm{d}\mathcal{L}(\Delta x, \Delta y)}{\mathrm{d}t} \right\rangle &= 2\frac{f_{\mathrm{rev}} \sum_i N_{\mathrm{Y};i} N_{\mathrm{B};i}}{\sqrt{(2\pi)^3}\,\Sigma_x\,\Sigma_y\,\Sigma_z} \times \\ &\times \int \sqrt{\frac{\beta_x^*\beta_y^*}{\beta_x(z)\beta_y(z)}} \exp\left(-\frac{\Delta x^2 \beta_x^*}{2\Sigma_x^2 \beta_x(z)} - \frac{\Delta y^2 \beta_y^*}{2\Sigma_y^2 \beta_y(z)} - \frac{4z^2}{2\Sigma_z^2} \right) \mathrm{d}z\end{aligned} \tag{A.18}$$

Then, assuming $\beta_x^* = \beta_y^* = \beta^*$ one can integrate analytically[3] and get a more general version of Eq. (A.11):

$$\begin{aligned}\left\langle \frac{\mathrm{d}\mathcal{L}(0,0)}{\mathrm{d}t} \right\rangle &= 2\frac{f_{\mathrm{rev}} \sum_i N_{\mathrm{Y};i} N_{\mathrm{B};i}}{\sqrt{(2\pi)^3}\,\Sigma_x\,\Sigma_y\,\Sigma_z} \int \left(1 + \frac{z^2}{(\beta^*)^2}\right)^{-1} \exp\left(-\frac{4z^2}{2\Sigma_z^2}\right) \mathrm{d}z = \\ &= \frac{f_{\mathrm{rev}} \sum_i N_{\mathrm{Y};i} N_{\mathrm{B};i}}{2\pi\,\Sigma_x\,\Sigma_y}\, \mathcal{J}\left(\frac{\sqrt{2}\beta^*}{\Sigma_z}\right)\end{aligned} \tag{A.19}$$

[2]Reads "beta star", with no connection to STAR, the experiment.

[3]A more general case can be computed numerically.

where form factor $\mathcal{J}(x) \equiv \sqrt{\pi} x \exp(x^2)\,(1-\mathrm{erf}(x))$ (see Fig. A.1 for numerical values).

The new expression of luminosity integral replacing Eq. (A.12) is:

$$\begin{aligned}
&\int \left\langle \frac{\mathrm{d}\mathcal{L}(\Delta x, 0)}{\mathrm{d}t} \right\rangle \mathrm{d}(\Delta x) = \\
&= 2\frac{f_{\mathrm{rev}} \sum_i N_{\mathrm{Y};i} N_{\mathrm{B};i}}{\sqrt{(2\pi)^3}\,\Sigma_x\,\Sigma_y\,\Sigma_z} \int \sqrt{\frac{\beta_x^* \beta_y^*}{\beta_x(z)\beta_y(z)}} \exp\left(-\frac{\Delta x^2 \beta_x^*}{2\Sigma_x^2 \beta_x(z)} - \frac{4z^2}{2\Sigma_z^2} \right) \mathrm{d}(\Delta x)\,\mathrm{d}z = \\
&= 2\frac{f_{\mathrm{rev}} \sum_i N_{\mathrm{Y};i} N_{\mathrm{B};i}}{\sqrt{4\pi^2}\,\Sigma_y\,\Sigma_z} \int \left(1 + \frac{z^2}{(\beta_y^*)^2} \right)^{-\frac{1}{2}} \exp\left(-\frac{4z^2}{2\Sigma_z^2} \right) \mathrm{d}z = \\
&= \frac{f_{\mathrm{rev}} \sum_i N_{\mathrm{Y};i} N_{\mathrm{B};i}}{\sqrt{2\pi}\,\Sigma_y}\, \mathcal{F}\left(\frac{\sqrt{2}\beta_y^*}{\Sigma_z} \right)
\end{aligned} \tag{A.20}$$

where form factor $\mathcal{F}(x) \equiv \sqrt{\frac{1}{\pi}}\, x \exp\left(\frac{x^2}{2}\right) K_0\left(\frac{x^2}{2}\right)$ (see Fig. A.1 for numerical values) where K_0(z) is the modified Bessel function of the second kind.

Then the expressions equivalent to Eqs. (A.13) and (A.14) (again, assuming $\beta_x^* = \beta_y^* = \beta^*$) are:

$$\Sigma_x = \frac{\mathcal{J}\left(\sqrt{2}\beta^*/\Sigma_z\right)}{\mathcal{F}\left(\sqrt{2}\beta^*/\Sigma_z\right)} \times \frac{\int \left\langle \frac{\mathrm{d}\mathcal{L}(\Delta x, 0)}{\mathrm{d}t} \right\rangle \mathrm{d}(\Delta x)}{\sqrt{2\pi} \left\langle \frac{\mathrm{d}\mathcal{L}(0,0)}{\mathrm{d}t} \right\rangle} \tag{A.21}$$

$$\Sigma_y = \frac{\mathcal{J}\left(\sqrt{2}\beta^*/\Sigma_z\right)}{\mathcal{F}\left(\sqrt{2}\beta^*/\Sigma_z\right)} \times \frac{\int \left\langle \frac{\mathrm{d}\mathcal{L}(0, \Delta y)}{\mathrm{d}t} \right\rangle \mathrm{d}(\Delta y)}{\sqrt{2\pi} \left\langle \frac{\mathrm{d}\mathcal{L}(0,0)}{\mathrm{d}t} \right\rangle} \tag{A.22}$$

If one instead uses Eqs. (A.13) and (A.14) as the definitions for Σ_x and Σ_y, then to calculate the luminosity one would have to multiply the RHS of Eqs. (A.11) and (A.12) by $\left[\mathcal{F}^2\left(\sqrt{2}\beta^*/\Sigma_z\right) / \mathcal{J}\left(\sqrt{2}\beta^*/\Sigma_z\right)\right]$, which is a correction to the Vernier scan luminosity measurement for the hourglass effect. The dependency of the different form factors introduced here on the ratio of β^* and Σ_z are presented in Fig. A.1. The hourglass effect shows up most strongly when β^* is less than Σ_z.

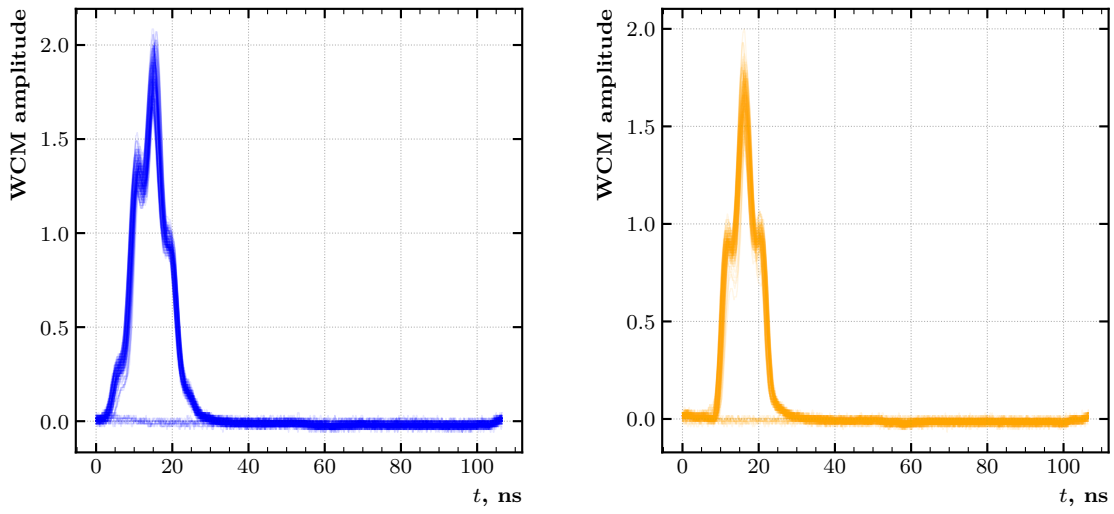
A.2 Luminosity in Run 12 at $\sqrt{s} = 200$ GeV

A.2.1 Vernier scans

During year 2012 there were three Vernier scans performed for proton-proton collisions at $\sqrt{s} = 200$ GeV at the STAR IP. They were performed near the end of RHIC fills 16444, 16470 and 16514. The combined transverse beam widths that were extracted in [79] from the Vernier scans are given in Table A.1.

RHIC Fill	Σ_x, μm	Σ_y, μm
16444	282 ± 9	277 ± 9
16470	310 ± 16	300 ± 12
16514	321 ± 25	312 ± 4

Table A.1: Values of combined widths of the beams measured along the x and y axes measured in Vernier scans. These values are read from plots in [79].



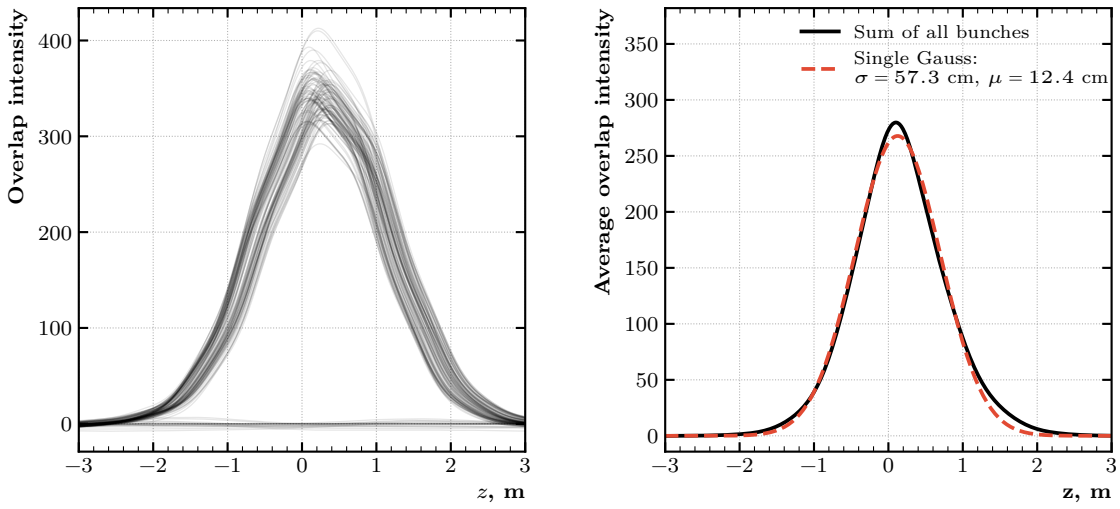
(a) Blue

(b) Yellow

Figure A.2: Raw WCM waveforms corresponding to longitudinal bunch profiles for each of the 120 bunches measured during the Vernier scan at the end of fill 16444. The multiple peak structures is typically characteristic of the beam shape before the RF rebucketing is performed in the beginning of a run, yet, for some reason, they are still seen here.

RHIC Fill	$\sigma_{z;\text{blue}}$, m	$\sigma_{z;\text{yellow}}$, m	$\Sigma_z/\sqrt{2}$, m
16444	1.30	1.13	1.22
16470	1.21	1.10	1.16
16514	0.99	0.71	0.86

Table A.2: Values of longitudinal widths of the beams measured using the WCM.



(a) (b)

Figure A.3: Convolutions of the raw WCM waveforms that were measured at end of fill 16444: (a) assuming no hourglass effect, showing each bunch pair, and (b) corrected for the hourglass effect, averaged over all bunch pairs (dashed curve shows a fit with a Gaussian curve). The global horizontal offset may not represent the actual overlap phase at any particular IP.

Hourglass effect

The measurements using the RHIC Wall Current Monitors (WCM) [103, 108] give longitudinal beam widths $\sigma_{z;1,2} \approx 1$ m (see Table A.2 for exact numbers). The nominal value of β^* at the STAR and PHENIX IPs was 0.85 m [109, 110]. The observed value of β^* and crossing angle $\Theta_{\times\text{-ing}}$ for the PHENIX IP was extracted from a Vernier scan performed at the PHENIX IP in fill 16470, giving $\beta^* \approx 80$ cm, which roughly matches the nominal value, and $\Theta_{\times\text{-ing}} \lessapprox 1$ mrad [111]. As seen in Fig. A.1, these numbers correspond to corrections to the luminosity values measured in the three Vernier scans from -4% to -1%.

To illustrate how the hourglass effect shapes the interaction diamond at the RHIC IPs, one can look at the raw waveform data from the WCMs (given in Fig. A.2). The intensity measured by a WCM located at a fixed point s_i along the beam trajectory is proportional to the beam density:

$$I_{\text{WCM};i}(t) \sim \int \rho_i(x, y, s_i, t)\, \mathrm{d}x\, \mathrm{d}y = \left.\frac{\mathrm{d}N}{\mathrm{d}z}\right|_{s=s_0-ct} \tag{A.23}$$

Similarly to the method in Eq. (A.18), one assumes the transverse density distribution in the drift space:

$$\rho_i(x, y, z, t) = \sqrt{\frac{\beta_x^* \beta_y^*}{\beta_x(z)\beta_y(z)}}\, \frac{1}{2\pi\sigma_{x;i}\sigma_{y;i}} \exp\left(-\frac{\Delta x^2 \beta_x^*}{2\sigma_x^2 \beta_x(z)} - \frac{\Delta y^2 \beta_y^*}{2\sigma_y^2 \beta_y(z)} \right) \left.\frac{\mathrm{d}N}{\mathrm{d}z}\right|_{s=z-ct} \tag{A.24}$$

To get the vertex distribution one omits the integration along z in Eq. (A.6) and substi-

tutes $\Delta x = \Delta y = 0$ to get:

$$\left\langle \frac{\mathrm{d}\mathcal{L}_{\text{int.}}}{\mathrm{d}z}(z) \right\rangle = 2c \int \rho_1(x,y,z,t)\, \rho_2(x,y,z,-t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \tag{A.25}$$

Substituting Eqs. (A.23) and (A.24) into Eq. (A.25) one gets the following equation describing the shape of the interaction vertex distribution:

$$\begin{aligned}\left\langle \frac{\mathrm{d}\mathcal{L}_{\text{int.}}}{\mathrm{d}z}(z) \right\rangle &\sim \frac{\beta_x^* \beta_y^*}{\beta_x(z)\beta_y(z)} \left[\int I_{\text{WCM};1}(z-ct)\, I_{\text{WCM};2}(z+ct)\, \mathrm{d}t \right] \times \\ &\times \left[\int \exp\left(-\frac{\Delta x^2 \Sigma_x^2 \beta_x^*}{2\sigma_{x;1}^2 \sigma_{x;2}^2 \beta_x(z)} - \frac{\Delta y^2 \Sigma_y^2 \beta_y^*}{2\sigma_{y;1}^2 \sigma_{y;2}^2 \beta_y(z)} \right) \mathrm{d}x\, \mathrm{d}y \right] \sim \\ &\sim \sqrt{\frac{\beta_x^* \beta_y^*}{\beta_x(z)\beta_y(z)}} \left[\int I_{\text{WCM};1}(z-ct)\, I_{\text{WCM};2}(z+ct)\, \mathrm{d}t \right]\end{aligned} \tag{A.26}$$

For the case $\beta_y^* = \beta_x^* = \beta$, and remembering Eq. (A.16) for $\beta(x)$ in the drift space, one gets:

$$\left\langle \frac{\mathrm{d}\mathcal{L}_{\text{int.}}}{\mathrm{d}z}(z) \right\rangle \sim \frac{1}{1 + \dfrac{z^2}{(\beta^*)^2}} \left[\int I_{\text{WCM};1}(z-ct)\, I_{\text{WCM};2}(z+ct)\, \mathrm{d}t \right] \tag{A.27}$$

The first multiplier is the hourglass correction to the vertex distribution and was previously seen in Eq. (A.19). The beam intensity convolution in the second multiplier can be calculated numerically from the measured longitudinal distributions. The resulting convolutions are presented in Fig. A.3a. The result averaged over all bunches and corrected using the nominal value of β^* is given in Fig. A.3b and is a bit wider than a usual vertex distribution seen at STAR, which may suggest a smaller real value of β^*.

Bunch intensities

The use of Eq. (A.11) or Eq. (A.12) requires knowledge of the numbers of protons in bunches that are needed to calculate the $\sum_i N_{1;i} N_{2;i}$. While the relative intensities of individual bunches can be readily measured by the WCMs, the absolute intensity normalization has to be calibrated [112] to the total beam current measured by the RHIC DC Current Transducer (DCCT) [103].

The convention is to enumerate bunches in the blue and yellow beams starting from the leading bunch with index 0 up to bunch 119. The bunches from 111 to 119 constitute the "abort gap" – they remain unfilled to facilitate a controlled beam dump. Each bunch of one beam always collides with a specific bunch of the other beam, however the pairing differs at the different IPs. At the PHENIX IP, which is at 8 o'clock, the bunches with equal IDs collide and the abort gaps align, but at the STAR IP, which is at 6 o'clock, the blue and yellow bunch IDs are shifted. The yellow beam rotating counterclockwise is thus 40 bunches late or 80 bunches ahead of the blue beam. Bunches 38 and 39 of the blue beam and bunches 78 and 79 of the yellow beam also remain unfilled; at the STAR IP, they align with the end of other beam's abort gaps.

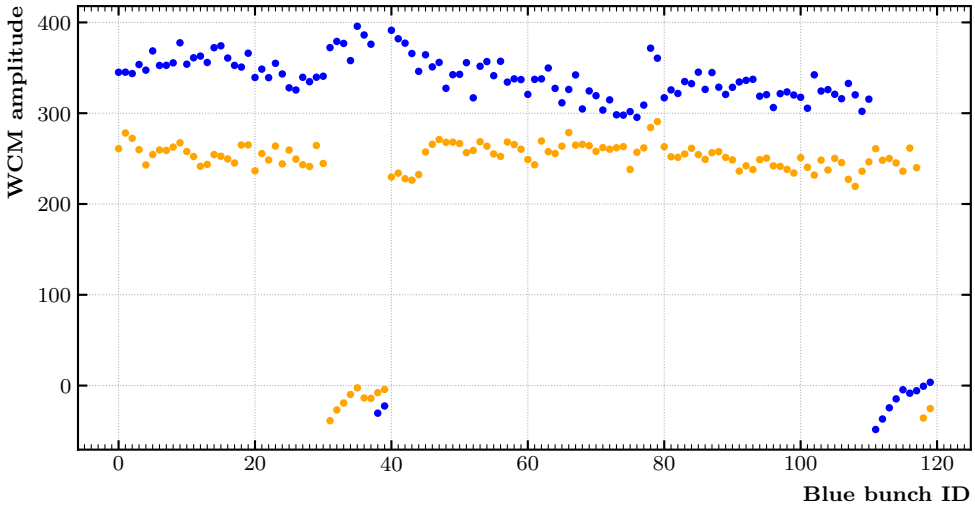


Figure A.4: Individual bunch intensities as measured by the WCM at the end of fill 16444. The blue bunches are enumerated according to the convention. The yellow bunches are aligned with the blue bunches that they collide with at the STAR IP.

This results in the following relation:

$$\sum_i N_{1;i} N_{2;i} \equiv \frac{N^{\mathrm{DCCT}}_{\mathrm{blue}} N^{\mathrm{DCCT}}_{\mathrm{yellow}}}{\left(\sum_{i=0}^{119} N^{\mathrm{WCM}}_{\mathrm{blue};i}\right)\left(\sum_{i=0}^{119} N^{\mathrm{WCM}}_{\mathrm{yellow};i}\right)} \sum_{i=0}^{119} N^{\mathrm{WCM}}_{\mathrm{blue};i} N^{\mathrm{WCM}}_{\mathrm{yellow};(i+80 \bmod 120)} \tag{A.28}$$

ZDC cross section

The ZDC coincidence cross section for Run 12 at $\sqrt{s} = 200$ GeV was determined from the data from the three Vernier scans to be $\sigma_{\mathrm{ZDC}} = 0.27$ mb [79].

A.2.2 Luminosity in the physics runs

The luminosity in a particular run is then given by:

$$\mathcal{L}_{\mathrm{int.}} = \frac{1}{\sigma_{\mathrm{ZDC}}} \int f_{\mathrm{ZDC}} \, \mathrm{d}t \tag{A.29}$$

The integrated luminosity corresponding to individual runs is shown in Fig. A.5. The total integrated luminosity used in this analysis amounts to 17 pb^{-1}.

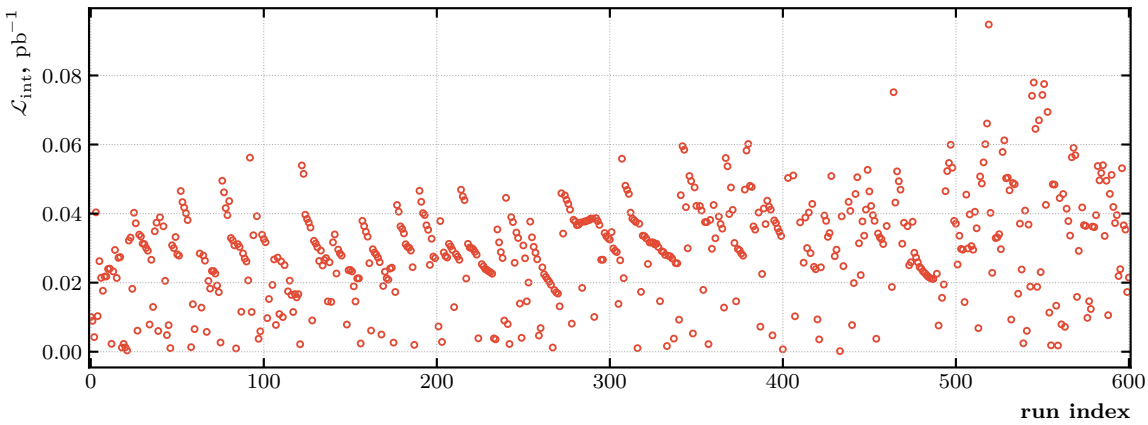


Figure A.5: Integrated luminosity for individual STAR runs.

Appendix B

Combining independent measurements

Suppose we have a set of measurements y_i along with their uncertainties $\delta\{y_i\}$ that are not correlated with each other. What is the best way to average them? We would like to minimize the uncertainty $\delta\{y\}$ of the average $y \equiv \sum_i \omega_i y_i$. The question is then what set of weights ω_i such that $\sum_i \omega_i = 1$ provides the optimal averaging.

To answer this question, let's follow the Lagrangian multipliers strategy. We start by writing down the Lagrangian function

$$\mathcal{L}(\omega_i, \lambda) = \delta\{\sum_i \omega_i y_i\} - \lambda(\sum_i \omega_i - 1) \tag{B.1}$$

The constrained minimum will then be a stationary point of $\mathcal{L}$, which satisfies the following system of equations

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \omega_j} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = 0 \end{cases} \tag{B.2}$$

Substituting the expressions for the derivatives we get

$$\begin{cases} \dfrac{\omega_j (\delta y_j)^2}{\delta\{\sum_i \omega_i y_i\}} - \lambda = 0 \\ \sum_i \omega_i - 1 = 0 \end{cases} \tag{B.3}$$

Expressing ω_j from the first equation in the system

$$\omega_j = \lambda\, \delta\{\sum_i \omega_i y_i\}\, (\delta y_j)^{-2} \tag{B.4}$$

And substituting into the second equation

$$\lambda\, \delta\{\sum_i \omega_i y_i\}\, \sum_k (\delta y_k)^{-2} = 1 \tag{B.5}$$

We arrive at the solution

$$\begin{cases} \lambda = \dfrac{1}{\delta\{\sum_i \omega_i y_i\} \left(\sum_k (\delta y_k)^{-2}\right)} \\ \omega_j = \dfrac{(\delta y_j)^{-2}}{\sum_i (\delta y_i)^{-2}} \end{cases} \tag{B.6}$$

Appendix C

Trigger promotion and demotion

In the formalism we will be using the following notation: For a predicate $\mathcal{P}$ over jets at the detector level, $\Theta(\mathcal{P}) = 1$ if $\mathcal{P}$ is true, and $\Theta(\mathcal{P}) = 0$ if $\mathcal{P}$ is false. Then the number of jets (e.g. in a given detector jet p_T bin) satisfying the condition $\mathcal{P}$ will be $N(\mathcal{P}) = \sum_{\text{jets}} \Theta(\mathcal{P})$. The "jets" subscript will be omitted. The logical conjunction "$\wedge$" (the "and" operator) is expressed as $\Theta(\mathcal{P} \wedge C) = \Theta(\mathcal{P})\,\Theta(C)$. The logical negation "$\overline{\ }$" (the "not" operator) is expressed as $\Theta(\overline{\mathcal{P}}) = 1 - \Theta(\mathcal{P})$.

We will assume that for each jet, the following relationships are true:

1. JP1 should fire $\Rightarrow$ JP0 should fire
   $\Theta(\text{JP0 should fire})\,\Theta(\text{JP1 should fire}) = \Theta(\text{JP1 should fire})$

2. JP2 should fire $\Rightarrow$ JP1 should fire
   $\Theta(\text{JP1 should fire})\,\Theta(\text{JP2 should fire}) = \Theta(\text{JP2 should fire})$

3. JP2 fired $\Leftrightarrow$ JP2 should fire
   $\Theta(\text{JP2 fired}) = \Theta(\text{JP2 fired})\,\Theta(\text{JP2 should fire}) = \Theta(\text{JP2 should fire})$

4. $\forall x \in \{0, 1, 2\}$:
   $\sum \Theta(\text{JPx fired})\,\Theta(\text{JPx should fire}) = \dfrac{1}{\text{PS(JPx)}} \sum \Theta(\text{JPx should fire})$

The last two assumptions in this list are not true for every jet, but we will for now neglect those jets that don't obey those rules. In the end, a consistency check will be done to ensure that the effect of those is small.

C.1 Trigger promotion

We will select the following three disjoint subsets of our real data:

- JP0 fired $\wedge$ $\overline{\text{JP1 should fire}}$ $\wedge$ JP0 should fire $\wedge$ other cuts
- (JP0 fired $\vee$ JP1 fired) $\wedge$ $\overline{\text{JP2 should fire}}$ $\wedge$ JP1 should fire $\wedge$ other cuts
- (JP0 fired $\vee$ JP1 fired $\vee$ JP2 fired) $\wedge$ JP2 should fire $\wedge$ other cuts

We will also consider three disjoint subsets of our embedding:

- $\overline{\text{JP1 should fire}} \wedge \text{JP0 should fire} \wedge \text{other cuts}$
- $\overline{\text{JP2 should fire}} \wedge \text{JP1 should fire} \wedge \text{other cuts}$
- $\text{JP2 should fire} \wedge \text{other cuts}$

$$
\begin{aligned}
&N(\text{JP0 fired} \wedge \overline{\text{JP1 should fire}} \wedge \text{JP0 should fire} \wedge \text{other cuts}) = \\
&\quad = \sum \Theta(\text{JP0 fired})\, \Theta(\overline{\text{JP1 should fire}})\, \Theta(\text{JP0 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP0)}} \sum \Theta(\overline{\text{JP1 should fire}})\, \Theta(\text{JP0 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP0)}}\, N(\overline{\text{JP1 should fire}} \wedge \text{JP0 should fire} \wedge \text{other cuts})
\end{aligned}
\tag{C.1}
$$

$$
\begin{aligned}
&N((\text{JP0 fired} \vee \text{JP1 fired}) \wedge \text{JP1 should fire} \wedge \overline{\text{JP2 should fire}} \wedge \text{other cuts}) = \\
&\quad = \sum \Theta(\text{JP0 fired} \vee \text{JP1 fired})\, \Theta(\overline{\text{JP2 should fire}})\, \Theta(\text{JP1 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \sum \left[\Theta(\text{JP0 fired}) + \Theta(\text{JP1 fired}) - \Theta(\text{JP0 fired})\, \Theta(\text{JP1 fired})\right] \\
&\qquad \Theta(\overline{\text{JP2 should fire}})\, \Theta(\text{JP1 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \left(\frac{1}{\text{PS(JP0)}} + \frac{1}{\text{PS(JP1)}} - \frac{1}{\text{PS(JP0)PS(JP1)}}\right) \\
&\qquad \sum \Theta(\overline{\text{JP2 should fire}})\, \Theta(\text{JP1 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \left(\frac{1}{\text{PS(JP0)}} + \frac{1}{\text{PS(JP1)}} - \frac{1}{\text{PS(JP0)PS(JP1)}}\right) \\
&\qquad N(\overline{\text{JP2 should fire}} \wedge \text{JP1 should fire} \wedge \text{other cuts})
\end{aligned}
\tag{C.2}
$$

$$
\begin{aligned}
&N((\text{JP0 fired} \vee \text{JP1 fired} \vee \text{JP2 fired}) \wedge \text{JP2 should fire} \wedge \text{other cuts}) = \\
&= \sum \Theta(\text{JP0 fired} \vee \text{JP1 fired} \vee \text{JP2 fired})\, \Theta(\text{JP2 should fire})\, \Theta(\text{other cuts}) = \\
&= \sum \big[\Theta(\text{JP0 fired}) + \Theta(\text{JP1 fired}) + \Theta(\text{JP2 fired}) \\
&\quad - \Theta(\text{JP0 fired})\Theta(\text{JP1 fired}) - \Theta(\text{JP1 fired})\Theta(\text{JP2 fired}) - \Theta(\text{JP2 fired})\Theta(\text{JP0 fired}) \\
&\quad + \Theta(\text{JP0 fired})\Theta(\text{JP1 fired})\Theta(\text{JP2 fired})\big]\, \Theta(\text{JP2 should fire})\, \Theta(\text{other cuts}) = \\
&= \sum \left[\frac{1}{\text{PS(JP0)}} + \frac{1}{\text{PS(JP1)}} + 1 \right. \\
&\quad - \frac{1}{\text{PS(JP0)}} \frac{1}{\text{PS(JP1)}} - \frac{1}{\text{PS(JP1)}} - \frac{1}{\text{PS(JP0)}} \\
&\quad \left. + \frac{1}{\text{PS(JP0)}} \frac{1}{\text{PS(JP1)}} \right]\, \Theta(\text{JP2 should fire})\, \Theta(\text{other cuts}) = \\
&= \sum \Theta(\text{JP2 should fire})\, \Theta(\text{other cuts}) = \\
&= N(\text{JP2 should fire} \wedge \text{other cuts})
\end{aligned}
\tag{C.3}
$$

C.2 Trigger demotion

Data categories:

- JP0 fired $\wedge$ JP0 should fire $\wedge$ other cuts
- $\overline{\text{JP0 fired}} \wedge$ JP1 fired $\wedge$ JP1 should fire $\wedge$ other cuts
- $\overline{\text{JP0 fired}} \wedge \overline{\text{JP1 fired}} \wedge$ JP2 fired $\wedge$ JP2 should fire $\wedge$ other cuts

Embedding categories:

- JP0 should fire $\wedge$ other cuts
- JP1 should fire $\wedge$ other cuts
- JP2 should fire $\wedge$ other cuts

Notice how in this case the embedding categories are overlapping, which uses the simulated sample statistics more effectively.

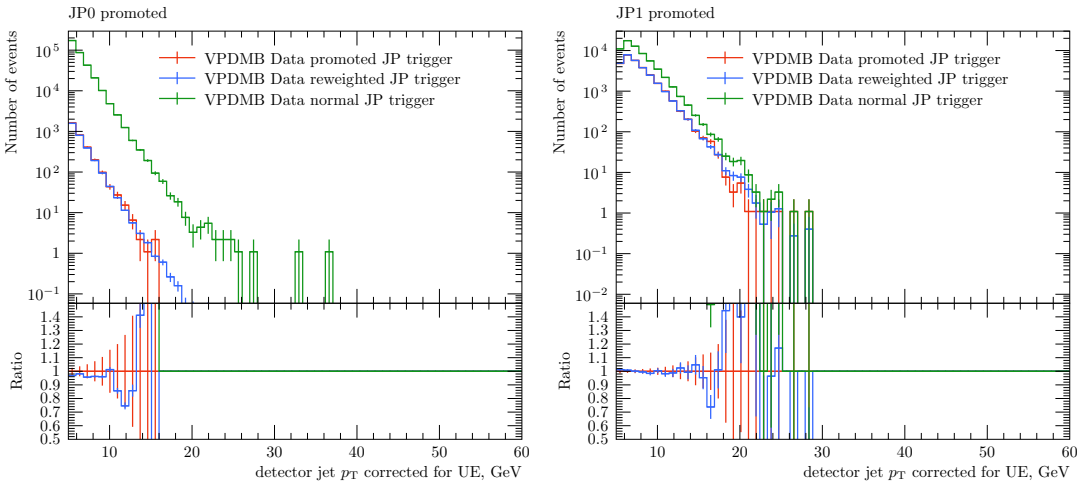
$$
\begin{aligned}
&N(\text{JP0 fired} \wedge \text{JP0 should fire} \wedge \text{other cuts}) = \\
&\quad = \sum \Theta(\text{JP0 fired})\, \Theta(\text{JP0 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP0)}} \sum \Theta(\text{JP0 should fire})\, \Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP0)}} N(\text{JP0 should fire} \wedge \text{other cuts})
\end{aligned}
\tag{C.4}
$$

$$
\begin{aligned}
&N(\overline{\text{JP0 fired}} \wedge \text{JP1 fired} \wedge \text{JP1 should fire} \wedge \text{other cuts}) = \\
&\quad = \sum \left[1 - \Theta(\text{JP0 fired})\Theta(\text{JP0 should fire})\right] \\
&\qquad \Theta(\text{JP1 fired})\,\Theta(\text{JP1 should fire})\,\Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP1)}} \sum \left(1 - \frac{\Theta(\text{JP0 should fire})}{\text{PS(JP0)}}\right) \Theta(\text{JP1 should fire})\,\Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP1)}} \left(1 - \frac{1}{\text{PS(JP0)}}\right) \sum \Theta(\text{JP1 should fire})\,\Theta(\text{other cuts}) = \\
&\quad = \frac{1}{\text{PS(JP1)}} \left(1 - \frac{1}{\text{PS(JP0)}}\right) N(\text{JP1 should fire} \wedge \text{other cuts})
\end{aligned}
\tag{C.5}
$$

$$
\begin{aligned}
&N(\overline{\text{JP0 fired}} \wedge \overline{\text{JP1 fired}} \wedge \text{JP2 fired} \wedge \text{JP2 should fire} \wedge \text{other cuts}) = \\
&\quad = \sum \left[1 - \Theta(\text{JP0 fired})\right]\left[1 - \Theta(\text{JP1 fired})\right] \\
&\qquad \Theta(\text{JP2 fired})\,\Theta(\text{JP2 should fire})\,\Theta(\text{other cuts}) = \\
&\quad = \sum \left(1 - \frac{1}{\text{PS(JP0)}}\right)\left(1 - \frac{1}{\text{PS(JP1)}}\right) \Theta(\text{JP2 should fire})\,\Theta(\text{other cuts}) = \\
&\quad = \left(1 - \frac{1}{\text{PS(JP0)}}\right)\left(1 - \frac{1}{\text{PS(JP1)}}\right) \sum \Theta(\text{JP2 should fire})\,\Theta(\text{other cuts}) = \\
&\quad = \left(1 - \frac{1}{\text{PS(JP0)}}\right)\left(1 - \frac{1}{\text{PS(JP1)}}\right) N(\text{JP2 should fire} \wedge \text{other cuts})
\end{aligned}
\tag{C.6}
$$

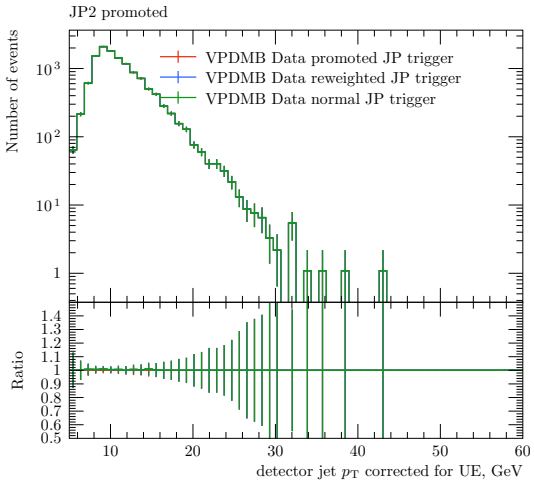
C.3 Consistency check

It was mentioned before that the assumptions used in the derivation are not exact facts. However, the nice thing about the relationships Eqs. (C.1) to (C.3) and Eqs. (C.4) to (C.6) is that they are valid not only for comparison between the data and embedding, but also for any sufficiently unbiased sample. For example, the VPDMB data sample can be evaluated using both the LHS and RHS of the both relationships. The comparison shown in Fig. C.1 and Fig. C.2 demonstrates a decent agreement.



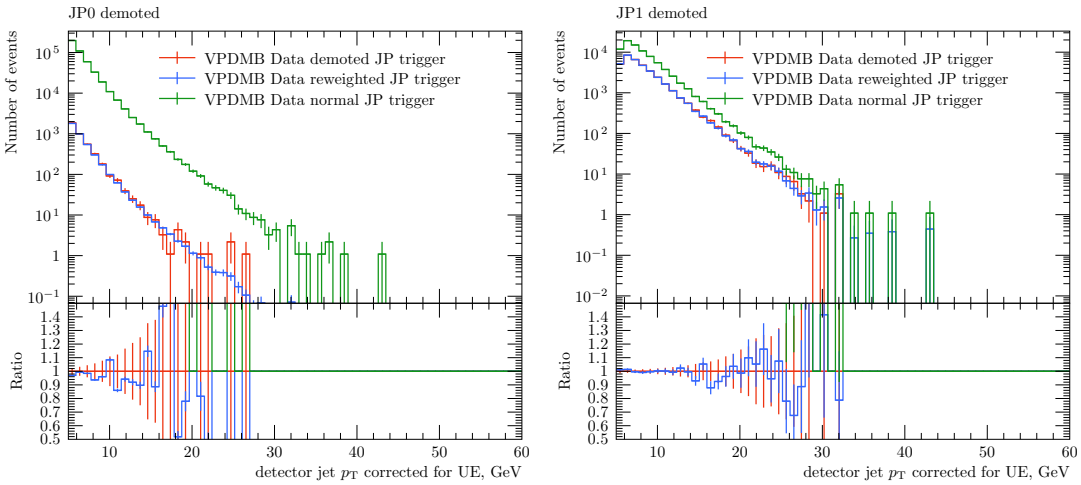
(a) JP0 promoted

(b) JP1 promoted



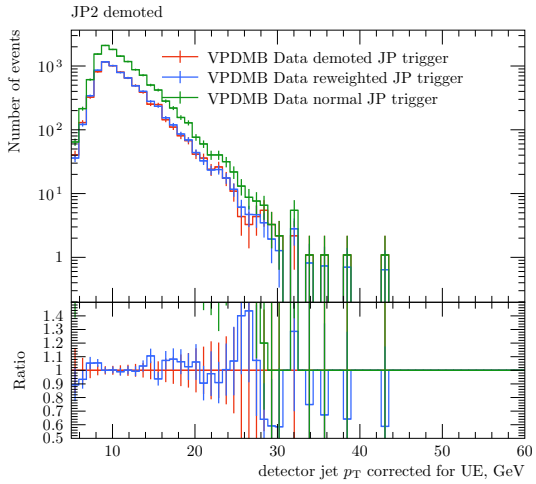
(c) JP2 promoted

Figure C.1: The jet yields for events from the data and embedding trigger promotion categories selected from the VPDMB data. The red line corresponds to the LHS of the Eqs. (C.1) to (C.3) and the blue line corresponds to the RHS.



(a) JP0 demoted

(b) JP1 demoted



(c) JP2 demoted

Figure C.2: The jet yields for events from the data and embedding trigger demotion categories selected from the VPDMB data. The red line corresponds to the LHS of the Eqs. (C.1) to (C.3) and the blue line corresponds to the RHS.

Appendix E

EMC systematic uncertainty formula

A standard formula for evaluating the BEMC uncertainty on the measured jet p_{T} that was used for previous jet spin asymmetry analyses in STAR looks like [113, 114]:

$$\frac{\delta\{p_{\mathrm{T}}^{\mathrm{jet}}\}}{p_{\mathrm{T}}^{\mathrm{jet}}} = \sqrt{R_{\mathrm{T}}^2 \left(\frac{\delta\{E_{\mathrm{T}}^{\mathrm{tower}}\}}{E_{\mathrm{T}}^{\mathrm{tower}}}\right)^2 + (1-R_{\mathrm{T}})^2 \left(\frac{\delta\{E_{\mathrm{T}}^{\mathrm{had.}}\}}{E_{\mathrm{T}}^{\mathrm{had.}}}\right)^2} \qquad \text{(6.20 revisited)}$$

where

$$\frac{\delta\{E_{\mathrm{T}}^{\mathrm{tower}}\}}{E_{\mathrm{T}}^{\mathrm{tower}}} \equiv \sqrt{\left(\frac{\delta\{\text{gain}\}}{\text{gain}}\right)^2 + \left(\frac{\delta\{\text{tower eff.}\}}{\text{tower eff.}}\right)^2} \qquad \text{(6.21 revisited)}$$

and

$$\frac{\delta\{E_{\mathrm{T}}^{\mathrm{had.}}\}}{E_{\mathrm{T}}^{\mathrm{had.}}} \equiv \left(\frac{1}{\epsilon_{\mathrm{trk}}} S_{\mathrm{hadron}} - p_{\mathrm{proj}}\right) \times f_{\mathrm{had.\,resp.}} \times \frac{\delta\{f_{\mathrm{had.\,resp.}}\}}{f_{\mathrm{had.\,resp.}}} \qquad \text{(6.22 revisited)}$$

Equation (6.20) is not an entirely obvious one, especially the term defined in Eq. (6.22). A good way to get an understanding for it is to try to derive it.

E.1 Proof

We start with a definition of measured $p_{\mathrm{T}}^{\mathrm{jet}}$ as a sum over its constituents:

$$p_{\mathrm{T}}^{\mathrm{jet}} = \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} + E_{\mathrm{T}}^{\mathrm{tower}} - \Delta E_{\mathrm{T}}^{\mathrm{sub.}} \qquad \text{(E.4)}$$

where the contribution of hadrons is given by the energy deposition of photons coming from hadron decay and deposition from long lived neutral and charged hadrons:

$$E_{\mathrm{T}}^{\mathrm{tower}} = \left(\sum_{\gamma} + \sum_{\mathrm{neut.\,had.}} + \sum_{\mathrm{chg.\,had.}}\right) E_{\mathrm{T}}^{\mathrm{tower}} \qquad \text{(E.5)}$$

The other term, $\Delta E_{\mathrm{T}}^{\mathrm{sub.}}$, represents the effect of the hadronic subtraction: towers that have tracks pointing to them have their energy reduced by the momentum of that track[1], and

[1] This method does not take into account the situation when the deposited energy gets split between several towers.

if the resulting tower energy is negative the tower is discarded. In practice, the charged hadrons will often deposit only a fraction of their energy in the calorimeter. We also neglect the contributions from electrons (that leave tracks and deposit most of their energy in the calorimeter via an electromagnetic shower), because those are relatively rare to come by in our events. We also ignore contributions from photons that accidentally hit the same tower as a track. So for the purposes of this derivation we just assume that if a tower has a track pointing to it, that tower will be fully subtracted:

$$\Delta E_{\mathrm{T}}^{\mathrm{sub.}} = \sum_{\mathrm{proj.\,tower}} E_{\mathrm{T}}^{\mathrm{tower}} \tag{E.6}$$

The variables in Eq. (6.22) would need to be evaluated using the following definitions:

$$f_{\mathrm{had.\,resp.}} \equiv \frac{\left\langle \sum_{\mathrm{proj.\,tower}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{proj.\,tower}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \tag{E.7}$$

to represent the fraction of energy that a charged hadron deposits in a calorimeter tower that its track projects to,

$$p_{\mathrm{proj}} \equiv \frac{\left\langle \sum_{\mathrm{proj.\,towers}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}{\left\langle \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \tag{E.8}$$

to represent the fraction of tracks that point to at least one EMC tower,

$$S_{\mathrm{hadron}} \equiv \frac{\left\langle \left(\sum_{\mathrm{neut.\,had.}} + \sum_{\mathrm{chg.\,had.}} \right) E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{chg.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle} \tag{E.9}$$

to parameterize an assumed scaling of neutral hadron energy with the charged hadron energy, and, finally,

$$\epsilon_{\mathrm{trk}} \equiv \frac{\left\langle \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}{\left\langle \sum_{\mathrm{chg.\,had.}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \tag{E.10}$$

to represent the tracking efficiency. In these four definitions, the $\langle ... \rangle$ denotes average of x in the events. One can notice that the definitions of the form

$$\frac{\left\langle \sum_i x \right\rangle_{\mathrm{events}}}{\left\langle \sum_j y \right\rangle_{\mathrm{events}}} \equiv \frac{\sum_{\mathrm{events}} \sum_i x}{\sum_{\mathrm{events}} \sum_j y} = \frac{\langle x \rangle_{\mathrm{events},i}}{\langle y \rangle_{\mathrm{events},j}} \tag{E.11}$$

that were used in Eqs. (E.7) to (E.10) deviate from a more common form

$$\left\langle \frac{x}{y} \right\rangle_{\mathrm{events}} \equiv \frac{1}{N_{\mathrm{events}}} \sum_{\mathrm{events}} \frac{x}{y}. \tag{E.12}$$

The definitions provided here allow to reduce the number of approximations that are needed for this proof. They are also similar to the ones that were used to determine the actual parameter values in [68].

Substituting Eqs. (E.5) and (E.6) into Eq. (E.4) one gets:

$$p_{\mathrm{T}}^{\mathrm{jet}} = \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} + \left(\sum_{\gamma} + \sum_{\mathrm{neut.\,had.}} + \sum_{\mathrm{chg.\,had.}} - \sum_{\mathrm{proj.\,tower}} \right) E_{\mathrm{T}}^{\mathrm{tower}} \tag{E.13}$$

First, looking closer at the last term in the brackets

$$\begin{aligned} \sum_{\mathrm{proj.\,tower}} E_{\mathrm{T}}^{\mathrm{tower}} &\approx \frac{\left\langle \sum_{\mathrm{proj.\,towers}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle \left\langle \sum_{\mathrm{proj.\,towers}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}{\left\langle \sum_{\mathrm{proj.\,towers}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle \left\langle \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} = \\ &= f_{\mathrm{had.\,resp.}}\, p_{\mathrm{proj}} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \end{aligned} \tag{E.14}$$

for which the following assumption had to be made:

$$\sum_{\mathrm{proj.\,tower}} E_{\mathrm{T}}^{\mathrm{tower}} \approx \frac{\left\langle \sum_{\mathrm{proj.\,tower}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}}. \tag{E.15}$$

Second, looking at two other terms:

$$\begin{aligned} &\left(\sum_{\mathrm{neut.\,had.}} + \sum_{\mathrm{chg.\,had.}} \right) E_{\mathrm{T}}^{\mathrm{tower}} \approx \\ &\approx \frac{\left\langle \left(\sum_{\mathrm{neut.\,had.}} + \sum_{\mathrm{chg.\,had.}} \right) E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle \dfrac{\left\langle \sum_{\mathrm{proj.\,towers}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{proj.\,towers}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}}{\left\langle \sum_{\mathrm{chg.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle \dfrac{\left\langle \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}{\left\langle \sum_{\mathrm{chg.\,had.}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} = \\ &= S_{\mathrm{hadron}} \frac{f_{\mathrm{had.\,resp.}}}{\epsilon_{\mathrm{trk}}} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \end{aligned} \tag{E.16}$$

for which three assumptions were used:

$$\sum_{\mathrm{chg.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \approx \frac{\left\langle \sum_{\mathrm{chg.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}}, \tag{E.17}$$

$$\sum_{\mathrm{neut.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \approx (S_{\mathrm{hadron}} - 1) \sum_{\mathrm{chg.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \tag{E.18}$$

and

$$\frac{\left\langle \sum_{\mathrm{proj.\,towers}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{proj.\,towers}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle} \approx \frac{\left\langle \sum_{\mathrm{chg.\,had.}} E_{\mathrm{T}}^{\mathrm{tower}} \right\rangle}{\left\langle \sum_{\mathrm{chg.\,had.}} p_{\mathrm{T}}^{\mathrm{track}} \right\rangle}. \tag{E.19}$$

Substituting Eqs. (E.14) and (E.16) into Eq. (E.13) one arrives at an expression

$$p_{\mathrm{T}}^{\mathrm{jet}} \approx \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}} + \sum_{\gamma} E_{\mathrm{T}}^{\mathrm{tower}} + \left(\frac{S_{\mathrm{hadron}}}{\epsilon_{\mathrm{trk}}} - p_{\mathrm{proj}} \right) f_{\mathrm{had.\,resp.}} \sum_{\mathrm{all}} p_{\mathrm{T}}^{\mathrm{track}}. \tag{E.20}$$

Finally, differentiating $p_{\mathrm{T}}^{\mathrm{jet}}$ by $f_{\mathrm{had.\,resp.}}$ one sees the expression for the second term under the square root in Eq. (6.20):

$$\frac{\partial p_{\mathrm{T}}^{\mathrm{jet}}}{\partial f_{\mathrm{had.\,resp.}}} \delta\{f_{\mathrm{had.\,resp.}}\} = \left(\frac{S_{\mathrm{hadron}}}{\epsilon_{\mathrm{trk}}} - p_{\mathrm{proj}} \right) f_{\mathrm{had.\,resp.}} \frac{\delta\{f_{\mathrm{had.\,resp.}}\}}{f_{\mathrm{had.\,resp.}}} (1 - R_{\mathrm{T}})\, p_{\mathrm{T}}^{\mathrm{jet}} \tag{E.21}$$

E.2 Discussion

Discrepancy in values of $f_{\mathrm{had.\,resp.}}$ estimated from data and embedding need not necessarily result in uncertainty. One could imagine that it could be rectified by tuning parameters of the detector simulation responsible for the response to hadrons (e.g., lowering the hadronic transport cut [115] or switching from GCALOR to GHEISHA [116]). Another possibility is that it could be implemented as a correction to be applied to the simulation, although that may require to have information about particle-tower association that is not easily available in our current software framework.

Bibliography

[1] A. S. Stodolna et al., *Hydrogen Atoms under Magnification: Direct Observation of the Nodal Structure of Stark States*, Phys. Rev. Lett. **110**, 213001 (2013) (cit. on p. 2).

[2] P. D. Group et al., *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020**, 083C01, 10.1093/ptep/ptaa104 (2020), eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673740/rpp2020-vol2-2015-2092_18.pdf (cit. on p. 3).

[3] J. Adam et al. (STAR), *Results on total and elastic cross sections in proton–proton collisions at $\sqrt{s}$ = 200 GeV*, Phys. Lett. B **808**, 135663 (2020), arXiv:2003.12136 [hep-ex] (cit. on p. 4).

[4] R. Field, *The Underlying Event in Hadronic Collisions*, Annu. Rev. Nucl. Part. Sci. **62**, 453–483 (2012) (cit. on p. 4).

[5] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics **2006**, 026–026 (2006), arXiv:hep-ph/0603175 (cit. on pp. 4, 38, 43).

[6] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5**, 1–91 (1989), arXiv:hep-ph/0409313 (cit. on p. 5).

[7] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, *Light-Cone Parton Distribution Functions from Lattice QCD*, Phys. Rev. Lett. **121**, 10.1103/physrevlett.121.112001 (2018), arXiv:1811.01588 [hep-lat] (cit. on p. 7).

[8] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, J. High Energy Phys. **2008**, 063–063 (2008), arXiv:0802.1189 [hep-ph] (cit. on pp. 10, 28).

[9] T. Åkesson et al. (AFS), *Direct evidence for the emergence of jets in events triggered on large transverse energy in pp collisions at $\sqrt{s} = 63$ GeV*, Phys. Lett. B **118**, 185–192 (1982), eprint: http://cds.cern.ch/record/139814 (cit. on pp. 11, 13).

[10] M. Banner et al. (UA2), *Observation of very large transverse momentum jets at the CERN $\bar{p}p$ collider*, Phys. Lett. B **118**, 203–210 (1982), eprint: http://cds.cern.ch/record/139969 (cit. on pp. 12, 13).

[11] G. Arnison et al. (UA1), *Observation of jets in high transverse energy events at the CERN proton antiproton collider*, Phys. Lett. B **123**, 115–122 (1983), eprint: http://cds.cern.ch/record/141936 (cit. on pp. 12, 13).

[12] P. Seyboth et al., *PRELIMINARY LARGE P(T) CROSS-SECTIONS MEASURED WITH A 2 PI CALORIMETER TRIGGER*, in 15th Rencontres de Moriond: Part I, High-Energy Hadronic Interactions (1980), eprint: https://inspirehep.net/literature/160970 (cit. on p. 13).

[13] K. P. Pretzl, *A study of deep inelastic hadron hadron collisions with a large acceptance transverse energy trigger*, AIP Conference Proceedings **85**, 585–601 (1982) (cit. on p. 13).

[14] T. Åkesson et al. (AFS), *Further evidence for jet production from the charged particles produced in pp collisions at $\sqrt{s} = 63$ GeV triggered on large transverse energy*, Phys. Lett. B **118**, 193–198 (1982), eprint: http://cds.cern.ch/record/139815 (cit. on p. 13).

[15] B. Abbott et al. (D0), *The Inclusive Jet Cross Section in $\bar{p}p$ Collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **82**, 2451–2456 (1999), arXiv:hep-ex/9807018 (cit. on p. 13).

[16] B. Abbott et al. (D0), *Inclusive Jet Production in $p\bar{p}$ Collisions*, Phys. Rev. Lett. **86**, 1707–1712 (2001), arXiv:hep-ex/0011036 (cit. on p. 13).

[17] V. Abazov et al. (D0), *The inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV using the k_T algorithm*, Phys. Lett. B **525**, 211–218 (2002), arXiv:hep-ex/0109041 (cit. on p. 13).

[18] T. Aaltonen et al. (D0), *Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron $p\bar{p}$ Collider Using a Cone-Based Jet Algorithm*, Phys. Rev. D **78**, 10.1103/physrevd.78.052006 (2008), arXiv:0802.2400 [hep-ex] (cit. on p. 13).

[19] V. M. Abazov et al. (D0), *Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D **85**, 10.1103/physrevd.85.052006 (2012), arXiv:1110.3771 [hep-ex] (cit. on p. 13).

[20] A. Abulencia et al. (CDF), *Measurement of the Inclusive Jet Cross Section using the k_T algorithm in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **96**, 10.1103/physrevlett.96.122001 (2006), arXiv:hep-ex/0512062 (cit. on p. 13).

[21] B. I. Abelev et al. (STAR), *Longitudinal Double-Spin Asymmetry and Cross Section for Inclusive Jet Production in Polarized Proton Collisions at $\sqrt{s} = 200$ GeV*, Phys. Rev. Lett. **97**, 10.1103/physrevlett.97.252001 (2006), arXiv:hep-ex/0608030 (cit. on p. 13).

[22] A. Tumasyan et al. (CMS), *Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $$ \sqrt{s} $$= 13 TeV*, Journal of High Energy Physics **2022**, 142 (2022), arXiv:2111.10431 [hep-ex] (cit. on p. 14).

[23] H. Paukkunen and P. Zurita, *PDF reweighting in the Hessian matrix approach*, J. High Energy Phys. **2014**, 10.1007/jhep12(2014)100 (2014) (cit. on pp. 14, 102).

[24] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. D **93**, 033006 (2016), arXiv:1506.07443 [hep-ph] (cit. on p. 14).

[25] N. Collaboration, *NNPDF 3.1 Catalog of plots*, (2017) https://data.nnpdf.science/nnpdf31-gallery/specs5_each_report_report.html (cit. on p. 14).

[26] B. Abelev et al., *Charged jet cross sections and properties in proton-proton collisions ats=7 TeV*, Phys. Rev. D **91**, 10.1103/physrevd.91.112012 (2015), arXiv:1411.4969 [nucl-ex] (cit. on p. 15).

[27] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. C **72**, 1896 (2012), arXiv:1111.6097 [hep-ph] (cit. on p. 15).

[28] M. Harrison, S. Peggs, and T. Roser, *THE RHIC ACCELERATOR*, Annu. Rev. Nucl. Part. Sci. **52**, 425–469 (2002), eprint: https://doi.org/10.1146/annurev.nucl.52.050102.090650 (cit. on p. 17).

[29] *RHIC Configuration Manual*, (2006) (cit. on p. 17).

[30] A. Zelenski, J. Alessi, A. Kponou, and D. Raparia, *High-intensity polarized H- (PROTON), deuteron and 3HE++ ion source development at BNL*, EPAC 2008 - Contributions to the Proceedings (2008) (cit. on p. 17).

[31] F. Bergsma et al., *The STAR detector magnet subsystem*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **499**, The Relativistic Heavy Ion Collider Project: RHIC and its Detectors, 633–639 (2003) (cit. on p. 19).

[32] M. Anderson et al., *The STAR time projection chamber: a unique tool for studying high multiplicity events at RHIC*, Nucl. Instrum. Methods Phys. Res., Sect. A **499**, The Relativistic Heavy Ion Collider Project: RHIC and its Detectors, 659–678 (2003) (cit. on p. 19).

[33] M. Beddo et al. (STAR), *The STAR barrel electromagnetic calorimeter*, Nucl. Instrum. Meth. A **499**, 725–739 (2003) (cit. on p. 19).

[34] C. Allgower et al. (STAR), *The STAR endcap electromagnetic calorimeter*, Nucl. Instrum. Meth. A **499**, 740–750 (2003) (cit. on p. 20).

[35] C. Adler, A. Denisov, E. Garcia, M. Murray, H. Stroebele, and S. White, *The RHIC zero degree calorimeters*, Nucl. Instrum. Methods Phys. Res., Sect. A **470**, 488–499 (2001) (cit. on p. 20).

[36] Y.-F. Xu, J.-H. Chen, Y.-G. Ma, A.-H. Tang, Z.-B. Xu, and Y.-H. Zhu, *Physics performance of the STAR zero degree calorimeter at relativistic heavy ion collider*, Nucl. Sci. Tech. **27**, 126 (2016) (cit. on p. 20).

[37] W. J. Llope et al., *The STAR Vertex Position Detector*, Nucl. Instrum. Meth. A **759**, 23–28 (2014), arXiv:1403.6855 [physics.ins-det] (cit. on p. 20).

[38] E. Judd et al., *The evolution of the STAR Trigger System*, Nucl. Instrum. Methods Phys. Res., Sect. A **902**, 228–237 (2018) (cit. on p. 21).

[39] F. Bieser et al., *The STAR trigger*, Nucl. Instrum. Methods Phys. Res., Sect. A **499**, The Relativistic Heavy Ion Collider Project: RHIC and its Detectors, 766–777 (2003) (cit. on p. 21).

[40] E. J. J.M. Engelage C. Gagliardi, *The Cabling Scheme and DSM Algorithms for the EMC Trigger*, tech. rep. (2017) (cit. on p. 21).

[41] J. K. Adkins and J. L. Drachenberg, *Azimuthal Single-Spin Asymmetries of Charged Pions in Jets in $s = 200$ GeV $p\uparrow p$ Collisions at STAR*, International Journal of Modern Physics: Conference Series **40**, 1660040 (2016), eprint: https://doi.org/10.1142/S2010194516600405 (cit. on p. 23).

[42] J. K. Adkins, *Studying Transverse Momentum Dependent Distributions in Polarized Proton Collisions Via Azimuthal Single Spin Asymmetries of Charged Pions in Jets*, PhD thesis (University of Kentucky, 2017) (cit. on pp. 23, 27, 39, 90).

[43] J. Adam et al. (STAR), *J/ψ production cross section and its dependence on charged-particle multiplicity in $p+p$ collisions at $\sqrt{s} = 200$ GeV*, Phys. Lett. B **786**, 87–93 (2018), arXiv:1805.03745 [hep-ex] (cit. on p. 23).

[44] J. Adam et al. (STAR), *Underlying event measurements in $p+p$ collisions at $\sqrt{s} = 200$ GeV at RHIC*, Phys. Rev. D **101**, 052004 (2020), arXiv:1912.08187 [nucl-ex] (cit. on pp. 23, 39, 40).

[45] J. Adam et al. (STAR), *Measurement of groomed jet substructure observables in p+p collisions at $\sqrt{s}$ =200 GeV with STAR*, Phys. Lett. B **811**, 135846 (2020), arXiv:2003.02114 [hep-ex] (cit. on p. 23).

[46] J. Adam et al. (STAR), *Measurement of inclusive J/ψ polarization in $p+p$ collisions at $\sqrt{s}$ =200 GeV by the STAR experiment*, Phys. Rev. D **102**, 092009 (2020), arXiv:2007.04732 [hep-ex] (cit. on p. 23).

[47] M. Abdallah et al. (STAR), *Invariant Jet Mass Measurements in pp Collisions at $\sqrt{s} = 200$ GeV at RHIC*, Phys. Rev. D **104**, 052007 (2021), arXiv:2103.13286 [hep-ex] (cit. on p. 23).

[48] M. A. Lisa (STAR), *The STAR TPC Clusterfinder/Hitfinder*, tech. rep. (1996) (cit. on p. 26).

[49] G. Van Buren et al., *Correcting for distortions due to ionization in the STAR TPC*, Nucl. Instrum. Meth. A **566**, 22–25 (2006), arXiv:0512157 [physics] (cit. on p. 26).

[50] G. V. B. and, *Precision Calibration of the STAR TPC*, in Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications (2008) (cit. on p. 26).

[51] F. Böhmer et al., *Simulation of Space-Charge Effects in an Ungated GEM-based TPC*, Nucl. Instrum. Meth. A **719**, 101–108 (2013), arXiv:1209.0482 [physics.ins-det] (cit. on p. 26).

[52] I. Chakaberia (STAR), *Grid leak simulation for STAR TPC using GARFIELD*, tech. rep. (2018) (cit. on p. 26).

[53] G. Dellacasa et al. (ALICE Collaboration), *ALICE time projection chamber: Technical Design Report*, Technical design report. ALICE (CERN, Geneva, 2000) (cit. on p. 26).

[54] D. Smirnov, J. Lauret, V. Perevoztchikov, G. V. Buren, and J. Webb, *Vertex Reconstruction at STAR: Overview and Performance Evaluation*, J. Phys. Conf. Ser. **898**, 042058 (2017) (cit. on p. 27).

[55] M. Cormier, A. Pavlinov, M. Rykov, V. Rykov, and K. Shestermanov, *STAR Barrel Electromagnetic Calorimeter Absolute Calibration Using "Minimum Ionizing Particles" from Collisions at RHIC*, tech. rep. (2001) (cit. on p. 28).

[56] C. Gagliardi, *Jet Finding Techniques discussion forum*, (2010) https://www.star.bnl.gov/HyperNews-star/protected/get/jetfinding/992/1.html (cit. on p. 29).

[57] G. V. Buren, *TPC Pt and DCA resolution*, (2008) https://drupal.star.bnl.gov/STAR/subsys/tpc/perf/tpc-pt-and-dca-resolution (cit. on p. 29).

[58] B. Page, *Run 9 200GeV Dijet High Pt Tail Investigation*, (2011) https://drupal.star.bnl.gov/STAR/blog/pagebs/2011/jul/28/run-9-200gev-dijet-high-pt-tail-investigation (cit. on p. 29).

[59] J. Adams et al. (STAR), *Measurements of transverse energy distributions in* Au + Au *collisions at* $\sqrt{s_{NN}} = 200$ GeV, Physical Review C **70**, 10.1103/physrevc.70.054907 (2004), arXiv:0407003 [nucl-ex] (cit. on pp. 29, 90).

[60] Z. Chang, *Hadron Energy Deposition in BEMC Towers*, (2019) https://drupal.star.bnl.gov/STAR/blog/zchang/hadron-energy-deposition-bemc-towers (cit. on p. 29).

[61] D. D. Staszak, *Measurements of the Jet Cross Section and Spin Asymmetry ALL using Polarized Proton Beams at RHIC*, PhD thesis (University of California, Los Angeles, 2010) (cit. on p. 33).

[62] G. V. Buren, *TPC Calibration & Data-Readiness Tasks*, (2005) https://drupal.star.bnl.gov/STAR/comp/calib/tpc (cit. on p. 33).

[63] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, *GEANT3*, (1987) (cit. on p. 38).

[64] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Physical Review D **82**, 10.1103/physrevd.82.074018 (2010), arXiv:1005.3457 [hep-ph] (cit. on p. 39).

[65] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, Journal of High Energy Physics **2002**, 012–012 (2002) (cit. on p. 39).

[66] J. Adams et al. (STAR), *Pion, kaon, proton and anti-proton transverse momentum distributions from p+p and d+Au collisions at sNN=200 GeV*, Phys. Lett. B **616**, 8–16 (2005), arXiv:nucl-ex/0309012 (cit. on p. 39).

[67] G. Agakishiev et al. (STAR), *Identified hadron compositions in* $p + p$ *and Au+Au collisions at high transverse momenta at* $\sqrt{s_{\mathrm{NN}}} = 200$ *GeV*, Phys. Rev. Lett. **108**, 072302 (2012), arXiv:1110.0579 [nucl-ex] (cit. on p. 39).

[68] E. Aschenauer and Z. Chang, *Inclusive Jet Cross-section Measurement in* 510 *GeV pp Collisions at STAR*, tech. rep. (Physics Department, Brookhaven National Laboratory, 2021) (cit. on pp. 40, 90, 96, 131).

[69] R. Field, *R. D. Field CDF Activities: PYTHIA tune Set A*, (2002) http://www.phys.ufl.edu/~rfield/cdf/tunes/py_tuneA.html (cit. on p. 43).

[70] D. Kalinkin, *Tracking efficiency uncertainty for the STAR TPC*, tech. rep. (Indiana University, 2020) (cit. on pp. 49, 92).

[71] A. Höcker and V. Kartvelishvili, *SVD approach to data unfolding*, Nucl. Instrum. Methods Phys. Res., Sect. A **372**, 469–481 (1996), arXiv:9509307 [hep-ph] (cit. on pp. 62, 66, 67).

[72] C. Davidson-Pilon, "Bayesian methods for hackers: probabilistic programming and Bayesian inference", in (Addison-Wesley, New York, 2016) Chap. 6 (cit. on p. 65).

[73] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Methods Phys. Res., Sect. A **362**, 487–498 (1995), eprint: http://cds.cern.ch/record/265717 (cit. on p. 68).

[74] L. B. Lucy, *An iterative technique for the rectification of observed distributions*, The Astronomical Journal **79**, 745 (1974) (cit. on p. 68).

[75] W. H. Richardson, *Bayesian-Based Iterative Method of Image Restoration∗*, J. Opt. Soc. Am. **62**, 55 (1972) (cit. on p. 68).

[76] G. D'Agostini, *Improved iterative Bayesian unfolding*, (2010), arXiv:1010.0632 [physics.data-an] (cit. on p. 68).

[77] S. Banach, *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales*, Fundamenta Mathematicae **3**, 133–181 (1922) (cit. on p. 69).

[78] L. Huo, *In-Jet Tracking Efficiency Analysis for the STAR Time Projection Chamber in Polarized Proton-Proton Collisions at* $\sqrt{s} = 200$ *GeV*, MA thesis (Texas A&M University, 2012) (cit. on p. 90).

[79] Z. Chang, *Vernier Scan*, (2020) https://www.star.bnl.gov/~zchang/vernier2/index.html (cit. on pp. 96, 116, 117, 120).

[80] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, J. High Energy Phys. **2002**, 029–029 (2002), arXiv:0204244 [hep-ph] (cit. on p. 98).

[81] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, J. High Energy Phys. **2004**, 040–040 (2004), arXiv:hep-ph/0409146 (cit. on p. 98).

[82] N. Sato, J. F. Owens, and H. Prosper, *Bayesian Reweighting for Global Fits*, Phys. Rev. D **89**, 114020 (2014), arXiv:1310.1089 [hep-ph] (cit. on p. 102).

[83] S. Alekhin et al., *HERAFitter*, Eur. Phys. J. C **75**, 304 (2015), arXiv:1410.4412 [hep-ph] (cit. on p. 102).

[84] H. Abramowicz et al. (H1, ZEUS), *Combination of measurements of inclusive deep inelastic* $e^{\pm}p$ *scattering cross sections and QCD analysis of HERA data*, Eur. Phys. J. C **75**, 580 (2015), arXiv:1506.06042 [hep-ex] (cit. on pp. 102, 104–107).

[85] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. D **93**, 033006 (2016), arXiv:1506.07443 [hep-ph] (cit. on p. 102).

[86] 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**, 204 (2015), arXiv:1412.3989 [hep-ph] (cit. on p. 102).

[87] R. D. Ball et al. (NNPDF), *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77**, 663 (2017), arXiv:1706.00428 [hep-ph] (cit. on p. 102).

[88] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15**, 438–450 (1972) (cit. on p. 104).

[89] Y. 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) (cit. on p. 104).

[90] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126**, 298–318 (1977) (cit. on p. 104).

[91] M. Botje, *QCDNUM: Fast QCD Evolution and Convolution*, Comput. Phys. Commun. **182**, 490–532 (2011), arXiv:1005.1481 [hep-ph] (cit. on p. 104).

[92] F. James, *MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1*, (1994) (cit. on p. 104).

[93] J. Pumplin, D. R. Stump, and W. K. Tung, *Multivariate fitting and the error matrix in global analysis of data*, Phys. Rev. D **65**, 014011 (2001), arXiv:hep-ph/0008191 (cit. on p. 104).

[94] Z. Nagy, *Next-to-leading order calculation of three jet observables in hadron hadron collision*, Phys. Rev. D **68**, 094002 (2003), arXiv:hep-ph/0307268 (cit. on p. 105).

[95] S. Catani and M. H. Seymour, *A General algorithm for calculating jet cross-sections in NLO QCD*, Nucl. Phys. B **485**, [Erratum: Nucl.Phys.B 510, 503–504 (1998)], 291–419 (1997), arXiv:hep-ph/9605323 (cit. on p. 105).

[96] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober (fastNLO), *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, (2011), arXiv:1109.1310 [hep-ph] (cit. on p. 105).

[97] D. Kalinkin, *FastNLO interpolation tables for NLO pQCD predictions of Inclusive Jet Production in pp collisions at* $\sqrt{s} = 200$ *and* 510 *GeV*, version 5 (Zenodo, 2022) (cit. on p. 105).

[98] F. D. Aaron et al. (H1), *Inclusive Deep Inelastic Scattering at High* Q^2 *with Longitudinally Polarised Lepton Beams at HERA*, JHEP **09**, 061 (2012), arXiv:1206.7007 [hep-ex] (cit. on p. 106).

[99] T. Aaltonen et al. (CDF), *Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p anti-p Collider Using a Cone-Based Jet Algorithm*, Phys. Rev. D **78**, [Erratum: Phys.Rev.D 79, 119902 (2009)], 052006 (2008), arXiv:0807.2204 [hep-ex] (cit. on p. 107).

[100] V. M. Abazov et al. (D0), *Measurement of the inclusive jet cross-section in* $p\bar{p}$ *collisions at* $s^{(1/2)}$ *=1.96-TeV*, Phys. Rev. Lett. **101**, 062001 (2008), arXiv:0802.2400 [hep-ex] (cit. on p. 107).

[101] M. Furman, *The Moeller luminosity factor*, tech. rep. LBNL–53553, CBP Note-543 (2003) (cit. on p. 111).

[102] S. van der Meer, *Calibration of the effective beam height in the ISR*, tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31 (CERN, Geneva, 1968) (cit. on p. 112).

[103] M. Bai et al., *RHIC beam instrumentation*, Nucl. Instrum. Meth. A **499**, 372–387 (2003) (cit. on pp. 112, 118, 119).

[104] S. Ozaki and T. Roser, *Relativistic Heavy Ion Collider, its construction and upgrade*, Progress of Theoretical and Experimental Physics **2015**, 03A102, 10.1093/ptep/ptu093 (2015), eprint: https://academic.oup.com/ptep/article-pdf/2015/3/03A102/19301086/ptu093.pdf (cit. on p. 112).

[105] D. Cronin-Hennessy, A. Beretvas, and P. Derwent, *Luminosity monitoring and measurement at CDF*, Nucl. Instrum. Methods Phys. Res., Sect. A **443**, 37–50 (2000) (cit. on p. 113).

[106] J. Dunlop, *ZDC (simple, i.e. RICH/RHIC) scalers in Run 2013*, (2013) https://drupal.star.bnl.gov/STAR/blog/dunlop/2013/mar/15/zdc-simple-ie-richrhic-scalers-run-2013 (cit. on p. 113).

[107] J. Dunlop, *RICH, RHIC, TCIM scaler mappings*, (2015) https://drupal.star.bnl.gov/STAR/blog/dunlop/rich-rhic-tcim-scaler-mappings (cit. on p. 113).

[108] P. Cameron et al., *The RHIC wall current monitor system*, in Proceedings of the 1999 Particle Accelerator Conference (Cat. No.99CH36366), Vol. 3 (1999), 2146–2148 vol.3 (cit. on p. 118).

[109] A. Anders et al., *RHIC Polarized Proton Operation in Run 12*, Conf. Proc. C **1205201**, edited by V. Suller, 184–186 (2012), eprint: https://www.rhichome.bnl.gov/RHIC/Runs/RhicRun12pp.pdf (cit. on p. 118).

[110] W. Fischer, *RHIC Run Overview*, https://www.rhichome.bnl.gov/RHIC/Runs/index.html#Run-12 (cit. on p. 118).

[111] M. J. Beaumier, *Probing the Spin Structure of the Proton Using Polarized Proton-Proton Collisions and the Production of W Bosons*, PhD thesis (UC, Riverside (main), 2016) (cit. on p. 118).

[112] M. Wilinski, *RHIC Wall Current Monitor*, (2019) https://www.cadops.bnl.gov/Instrumentation/InstWiki/index.php/RHIC_Wall_Current_Monitor (cit. on p. 119).

[113] Z. Chang, *Inclusive Jet Longitudinal Double-spin Asymmetry A_{LL} Measurements in 510 GeV Polarized pp Collisions at STAR*, PhD thesis (Texas A-M, 2016), arXiv:1907.10368 [hep-ex] (cit. on p. 130).

[114] C. Gagliardi, *Jet energy uncertainty from BEMC response to hadrons*, (2018) https://drupal.star.bnl.gov/STAR/blog/gagliard/jet-energy-uncertainty-bemc-response-hadrons (cit. on p. 130).

[115] J. C. Webb, private communication, 2019 (cit. on p. 133).

[116] G. Mavromanolakis and D. Ward, *Comparisons of hadronic shower packages*, (2004), arXiv:physics/0409040 (cit. on p. 133).