IN-JET TRACKING EFFICIENCY ANALYSIS FOR THE STAR TIME PROJECTION CHAMBER IN POLARIZED PROTON-PROTON COLLISIONS AT $\sqrt{S} = 200$ GeV

A Thesis

by

LIAOYUAN HUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

October 2011

Major Subject: Physics

IN-JET TRACKING EFFICIENCY ANALYSIS FOR THE STAR TIME PROJECTION CHAMBER IN POLARIZED PROTON-PROTON COLLISIONS AT $\sqrt{S} = 200$ GeV

A Thesis

by

LIAOYUAN HUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Gagliardi, Carl A.
Committee Members,	Fries, Rainer J.
	Natowitz, Joseph B.
	Tribble, Robert E.
Head of Department,	Fry, Edward S.

October 2011

Major Subject: Physics

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

JETS IN PROTON-PROTON COLLISIONS AT RHIC

A. Proton spin and the measurement of Δg

Protons are fermions with a spin of $\frac{1}{2}$. The proton spin was first explained by the static constituent quark model, in which a proton consists of two up quarks and one down quark, and constituent quark spins contribute 100% of the proton spin. Relativistic constituent quark models, in which protons are considered as bound states of three confined constituent quarks, bring the contribution to proton spin from quark spins down to ∼60%. The remainder then arises from quark orbital motion. However, measurements from polarized Deep Inelastic Scattering (DIS) experiments, with polarized electrons or muons as probes, indicate that the quark contribution to the proton spin is ∼25% [1–4], which is well below the theoretical predictions. This difference is often cited as the "proton spin crisis".

Most recent understanding of the proton structure indicates that a proton contains three valence quarks with spin $\frac{1}{2}$, gluons with spin 1 which are the force carriers of strong interactions, and quark-anti-quark pairs (sea quarks) that are split from gluons and then annihilate into gluons. The quarks, anti-quarks and gluons are collectively referred to as partons. Spin contributions from them can be expressed as the proton-spin sum rule [5]:

$$\frac{1}{2} = \int_0^1 dx[\frac{1}{2}\sum_q(\Delta q + \Delta\overline{q})(x,\mu^2) + \Delta g(x,\mu^2)] + L_q(\mu^2) + L_g(\mu^2). \tag{1.1}$$

For a longitudinally polarized proton,

$$\Delta q(x,\mu^2) \equiv q_+^+(x,\mu^2) - q_+^-(x,\mu^2),$$
$$\Delta g(x,\mu^2) \equiv g_+^+(x,\mu^2) - g_+^-(x,\mu^2), \tag{1.2}$$

where superscripts refer to parton helicities and subscripts refer to the parent proton helicities, x is the Bjorken-x of the parton which specifies the fraction of the proton momentum that is carried by the parton, and μ is the factorization scale. L stands for the orbital angular momentum of quarks and gluons in the proton. DIS experiments with point-like electromagnetic probes have made detailed measurements of the distribution of $\sum_q(\Delta q + \Delta\overline{q})(x,\mu^2)$, but put little constraint on the gluon polarization distribution $\Delta g(x,\mu^2)$ [1–4].

Polarized proton-proton collisions use quarks and gluons as probes for Δg, through quark-gluon and gluon-gluon hard scattering. For the production of a final state particle, e.g., the production of a pion, $pp \to \pi X$, the cross-section over some generic final state phase space $d\mathcal{P}$ can be expressed as

$$\frac{d\sigma^{pp\to\pi X}}{d\mathcal{P}} = \sum_{f_1,f_2,f} \int dx_1 dx_2 dz f_1^p(x_1,\mu^2) f_2^p(x_2,\mu^2) \times \frac{d\hat{\sigma}^{f_1 f_2 \to f X'}}{d\mathcal{P}}(x_1 p_1, x_2 p_2, p_\pi/z, \mu) D_f^\pi(z,\mu^2), \tag{1.3}$$

where f_i^p is the probability density function of finding parton i in a proton with x[1], D_f^π is the hadronization fragmentation function, and $\frac{d\hat{\sigma}^{f_1 f_2 \to f X'}}{d\mathcal{P}}$ is the underlying hard-scattering cross-section. With different configurations of helicities of the colliding

[1]Therefore, $q \equiv f_q^p$, $g \equiv f_g^p$ in Eq. 1.2.

protons, one can define the polarized cross-section:

$$\begin{aligned}\frac{d\Delta\sigma^{pp\to\pi X}}{d\mathcal{P}} &\equiv \frac{1}{4}[\frac{d\sigma^{pp\to\pi X}_{++}}{d\mathcal{P}} - \frac{d\sigma^{pp\to\pi X}_{+-}}{d\mathcal{P}} - \frac{d\sigma^{pp\to\pi X}_{-+}}{d\mathcal{P}} + \frac{d\sigma^{pp\to\pi X}_{--}}{d\mathcal{P}}] \\ &= \sum_{f_1,f_2,f}\int dx_1 dx_2 dz \Delta f_1^p(x_1,\mu^2)\Delta f_2^p(x_2,\mu^2) \\ &\quad\times \frac{d\Delta\hat{\sigma}^{f_1f_2\to fX'}}{d\mathcal{P}}(x_1p_1, x_2p_2, p_\pi/z, \mu) D_f^\pi(z,\mu^2),\end{aligned} \tag{1.4}$$

where

$$\frac{d\Delta\hat{\sigma}^{f_1f_2\to fX'}}{d\mathcal{P}} \equiv \frac{1}{4}[\frac{d\hat{\sigma}^{f_1f_2\to fX'}_{++}}{d\mathcal{P}} - \frac{d\hat{\sigma}^{f_1f_2\to fX'}_{+-}}{d\mathcal{P}} - \frac{d\hat{\sigma}^{f_1f_2\to fX'}_{-+}}{d\mathcal{P}} + \frac{d\hat{\sigma}^{f_1f_2\to fX'}_{--}}{d\mathcal{P}}]. \tag{1.5}$$

The cross-sections of the underlying hard scattering processes can be calculated by perturbative QCD, while f_i^p and D_f^π can be measured in other experiments. If we can measure the polarized cross-section, we will be able to calculate the polarization distribution functions Δf_i^p.

Actually, it is more convenient to measure the longitudinal double-spin asymmetry:

$$\begin{aligned}A_{LL}^\pi &= \frac{d\Delta\sigma^{pp\to\pi X}/d\mathcal{P}}{d\sigma^{pp\to\pi X}/d\mathcal{P}} \\ &= \frac{1}{P_1P_2}\times\frac{(N'_{++}+N'_{--}) - R(N'_{+-}+N'_{-+})}{(N'_{++}+N'_{--}) + R(N'_{+-}+N'_{-+})},\end{aligned} \tag{1.6}$$

where P_1, P_2 are beam polarizations, N's are the event counts of the different spin configurations, and

$$R \equiv \frac{L_{++}+L_{--}}{L_{+-}+L_{-+}}, \tag{1.7}$$

which compensates the differences in beam luminosities. Therefore, once A_{LL} is measured, one will be able to put a constraint on Δg, together with the (calculable) unpolarized cross-sections and already measured distributions of Δq.

B. Jets and the measurement of jet asymmetry

In proton-proton collision experiments, jets are commonly defined as a collection of final-state particles that cluster into a small region in the coordinates of pseudo-rapidity−azimuth ($\eta - \phi$). Ideally, a jet consists of all the final-state particles from the hadronization of a single parton; therefore, the fragmentation function $D(z, \mu^2)$ is no longer required in the calculation for unpolarized and polarized cross-sections of jet production:

$$\frac{d\sigma^{pp\to jetX}}{d\mathcal{P}} = \sum_{f_1,f_2,f} \int dx_1 dx_2 f_1^p(x_1,\mu^2) f_2^p(x_2,\mu^2) \times \frac{d\hat{\sigma}^{f_1 f_2 \to fX'}}{d\mathcal{P}}(x_1 p_1, x_2 p_2, p_{jet}, \mu) \tag{1.8}$$

$$\frac{d\Delta\sigma^{pp\to jetX}}{d\mathcal{P}} = \sum_{f_1,f_2,f} \int dx_1 dx_2 \Delta f_1^p(x_1,\mu^2) \Delta f_2^p(x_2,\mu^2) \times \frac{d\Delta\hat{\sigma}^{f_1 f_2 \to fX'}}{d\mathcal{P}}(x_1 p_1, x_2 p_2, p_{jet}, \mu) \tag{1.9}$$

Next-to-Leading Order (NLO) perturbative QCD calculations indicate that quark-gluon (qg) and gluon-gluon(gg) hard scattering dominate the polarized cross-section in jet production at RHIC energies, as demonstrated in Fig. 1 [6]. Consequently, the single-jet double-spin asymmetry

$$A_{LL}^{jet} = \frac{d\Delta\sigma^{pp\to jetX}/d\mathcal{P}}{d\sigma^{pp\to jetX}/d\mathcal{P}} \tag{1.10}$$

serves as a preferred measurement to determine Δg, because of its sensitivity to the gluon distribution. One of the major detectors at the Relativistic Heavy-Ion Collider (RHIC), the Solenoidal Tracker At RHIC (STAR) [7], is particularly well suited to measure A_{LL}^{jet} because it has large and relatively uniform acceptance in the kinematic region where $d\Delta\sigma/d\sigma$ is large. By counting the number of single-jet events in different polarization configurations of colliding protons, A_{LL}^{jet} can be measured as functions of

jet kinematic variables, such as transverse momentum and pseudo-rapidity. In the end, one can eventually constrain the gluon polarization distribution by comparing the measured A_{LL}^{jet} against calculations from various theoretical models.

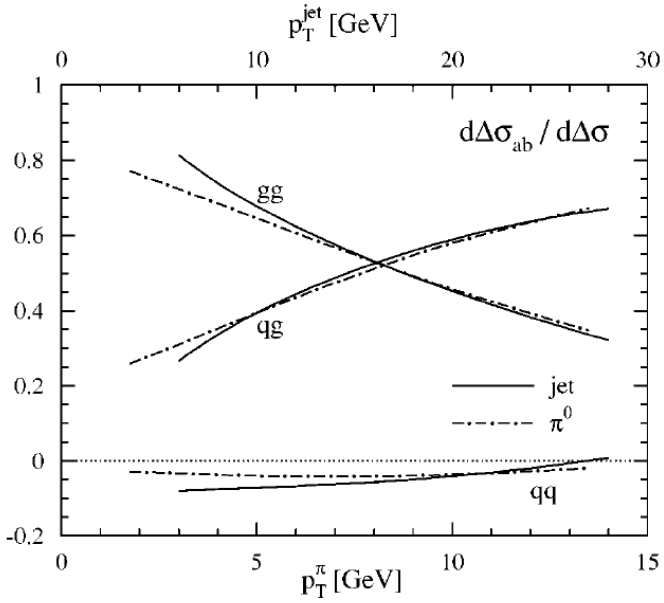


Fig. 1. Relative contributions from gg, qg, and qq scatterings to the NLO polarized cross-section for jet and inclusive-π^0 production at mid pseudo-rapidities at $\sqrt{s} = 200$ GeV.

C. Polarized Proton-Proton Collisions at RHIC

Located at the Brookhaven National Laboratory, RHIC is designed to perform experiments for both heavy-ion collisions with center-of-mass energy up to 200 GeV per nucleon-nucleon pair, and polarized proton-proton collisions with center-of-mass energy at 200 GeV and 500 GeV. Accelerated particles are carried in two separated rings travelling in opposite directions defined as the blue beam and the yellow beam. There are six interaction points located evenly on the RHIC rings, where the beams are steered and collide into each other. The two major detectors, STAR and PHENIX,

are located at the beam interaction points at six o'clock and eight o'clock of the 3.8-km-long RHIC beam ring. Figure 2 [5] demonstrates the scheme of the RHIC.

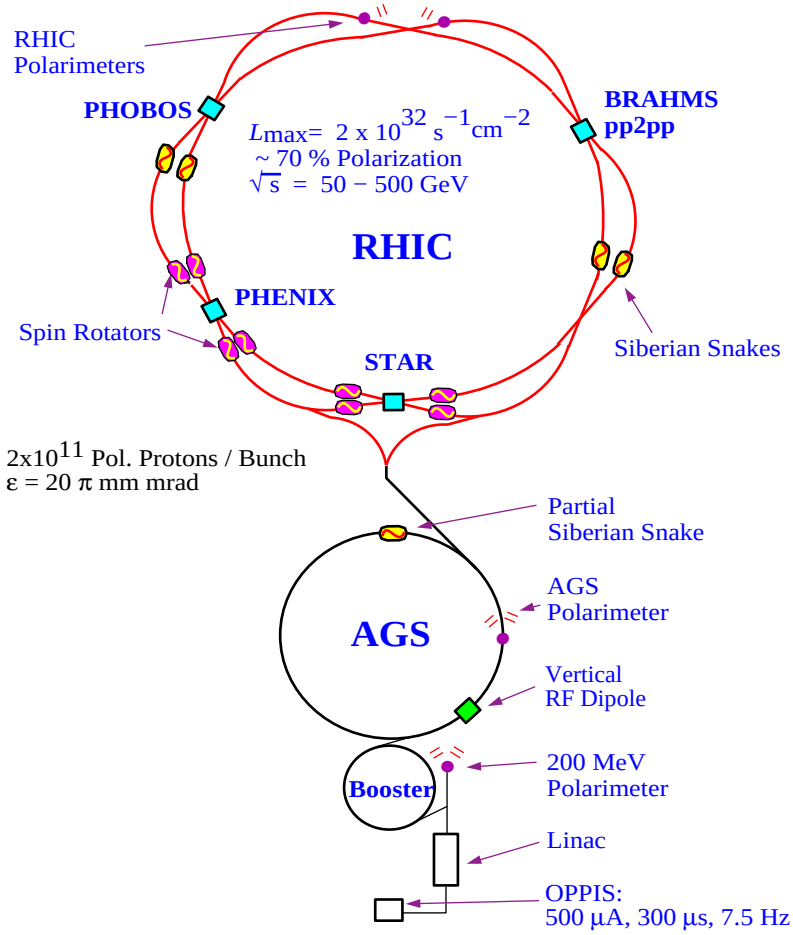


Fig. 2. Detectors and relevant devices for polarized proton-proton collisions.

Many devices are specifically designed to make RHIC the first accelerator in the world that is capable to perform polarized proton-proton collision experiments. Starting from a polarized proton source, the energy of the protons are brought up to 24 GeV through a series of accelerators before they are injected into the RHIC rings [8,9]. Protons are injected into each ring as "bunches" of protons which contain $1 \sim 2 \times 10^{11}$ protons each, with typically 110 bunches in total in each ring [10]. In the storage ring, accelerated protons have a stable spin state which is perpendicular

to the beam direction, i.e., the transverse spin state. Each ring is equipped with two magnets, referred as Siberian Snakes, that flip the spin state of the protons that are passing through them, in order to prevent the beam from depolarization. There are also four Spin Rotators located around the interaction points of both STAR and PHENIX, which can rotate the spin of protons parallel to the beam direction when the protons are entering the detectors, in order to perform longitudinally polarized proton collision experiments [11], and then rotate them back to the transverse stable state when they are leaving the interaction point. Various polarimeters are equipped to monitor beam polarization, with proton-Carbon polarimeters for fast beam polarization measurements, and a H-jet polarimeter for calibration.

The presented analysis uses the STAR data from the second longitudinal run period of 2006, with a center-of-mass energy of 200 GeV. This run period completed with an average polarization of 60% and an accumulated luminosity of $\sim 6\text{pb}^{-1}$.

CHAPTER II

BACKGROUND OF THE STAR DETECTOR AND THE TIME PROJECTION CHAMBER

A. Jet detection with STAR

1. The STAR detector and the Time Projection Chamber

STAR, the Solenoidal Tracker At RHIC, consists of several sub-systems that cover different rapidity ranges and functionality [7].

The Time Projection Chamber (TPC) [12], which is the innermost detector, provides tracking for charged particles such as charged pions, electrons, protons and etc. at mid-rapidity. The barrel side of TPC is covered by the Time of Flight detector (ToF), which is then covered by the Barrel Electro-Magnetic Calorimeter (BEMC), which provides energy measurement for electrons, direct gammas and short-life neutral particles such as neutral pions.

Other detectors include forward-rapidity detectors such as Forward TPC, End-cap Electro-Magnetic Calorimeter, Forward Pion Detector and Forward Meson Spectrometer, and detectors mainly used for triggering and luminosity monitoring, such as Beam-Beam Counter (BBC) and Zero Degree Calorimeter. Figure. 3 shows the layout of major sub-systems in the STAR detector.

Being the primary tracking device at mid-rapidity, the TPC is designed to provide tracking and momentum measurement of charged particles at high multiplicity. The TPC measures 4.2 meters long and 4 meters in outer diameter, which is equivalent to -1 to 1 in pseudo-rapidity with respect to detector center, and 1 meter in inner diameter. The chamber is filled with drift gas (P10) and divided into two chambers by the center membrane, which is set to a high voltage of 28 kV. The electric field

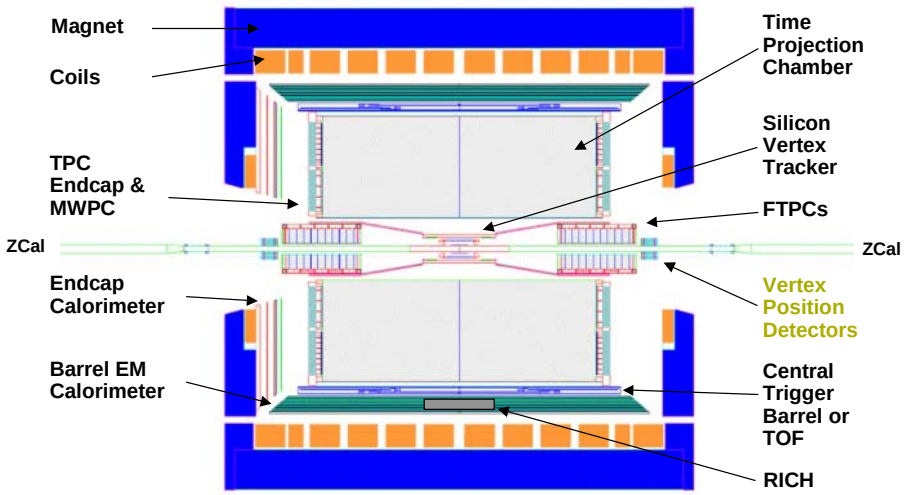


Fig. 3. The STAR detector

in the chamber is designed to be uniform and drops ∼135 V/cm toward the endcap anode planes, which are set at a voltage near ∼1 kV.

As charged particles travel through the TPC volume, secondary electrons will be excited along the track and start drifting toward the anode planes under the force of the uniform electric field. As they reach the TPC endcap, their position in x-y is read out by the anode planes (divided into 24 sectors on each endcap) and positions in z are determined by the drifting time in the chamber. Figure 4 demonstrates the structure of the TPC.

At the anode plane, the secondary electrons are first amplified by the Multi-Wire Proportional Chamber on the anode plane. As they reach the anode wires with a high voltage above 1 kV, the number of secondary electrons is amplified by the avalanche effect by a factor between 1000 to 3000. The excited electrons are absorbed by the anode wires, and the image charge of the positive ions left near anode wires is read out by the anode pads, which lie under the anode wires. Figure 5 shows the structure of

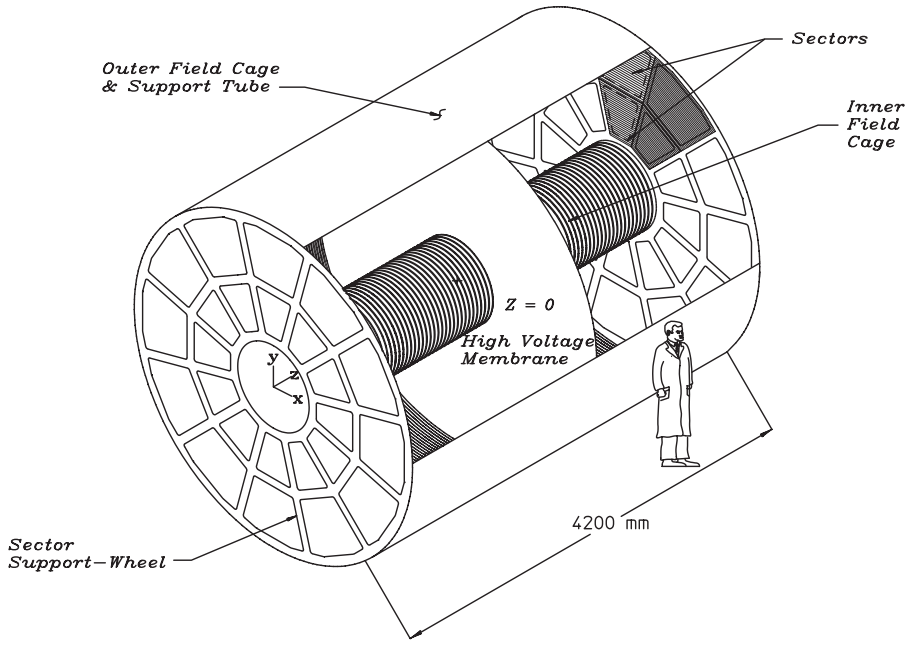


Fig. 4. The structure of the TPC.

the anode wires and read-out pads [13]. The choice of the directions of the anode wire and read-out pads provides a better position resolution in the azimuthal direction in order to improve track momentum resolution. The read-out pads are grouped in the azimuthal direction as "pad rows", and each TPC sector contains 45 of them.

2. Jet finding with TPC and BEMC

While the TPC provides tracking and momentum information for charged particles, the BEMC [14] measures energy and directional information of neutral short-life particles (by measuring gammas decayed from them), direct gammas and electrons/positrons. The list of charged and neutral particles is then fed to a jet finding algorithm for jet reconstruction.

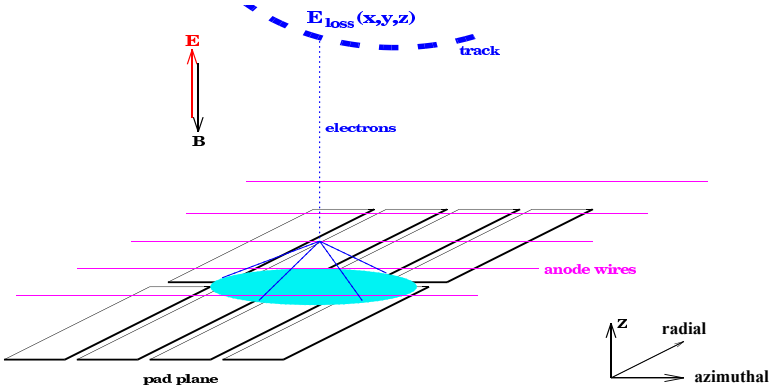


Fig. 5. The structure of anode wires and read-out pads

For each event, the STAR jet finding algorithm [15, 16] tries to find separated[1] jet cones that contain as much energy as possible, with a fixed jet cone radius. In this analysis, the mid-point cone algorithm [17, 18] with a radius of 0.7 in $\eta - \phi$ was used and the jet finding algorithm only considered TPC tracks with 12 or more fit points. Jet thrust axis is defined as the momentum weighted average direction of particles within a jet cone.

3. Jet event triggering

Jet events are triggered by the relatively fast BEMC detector. In RHIC Run 6 (Spring, 2006), most of the recorded jet events were triggered by the BEMC "jet patch" trigger. A "jet patch" is an area in the BEMC defined with a size of 1×1 in $\eta \times \phi$. With the size of $2 \times 2\pi$ in $\eta \times \phi$, the BEMC is divided into 12 jet patches with 6 each in the east and west halves. If the energy deposit in any of these patches is greater than the jet patch trigger threshold, the event will be triggered and recorded as a candidate jet event.

[1]Determined by the ratio of energy of overlapping particles to total jet energy. Often required to be smaller than 1/2.

4. The systematic uncertainty of TPC tracking efficiency in jets

In STAR, the TPC tracking efficiency within a data set is typically estimated through embedding, in which a simulated track is mixed with other signals of an event from real data. The mixed event is then treated as if it is from real data and reconstructed in the same way as in real data production. Then the tracking efficiency is estimated as the possibility of a simulated track being able to be matched with a reconstructed track.

Standard STAR analysis assumes ∼5% uncertainty on the TPC tracking efficiency, which is obtained in a study [12] which focused on identified particle measurements in heavy ion collisions, instead of jets in proton+proton collisions. The present study investigated the systematic uncertainty of the TPC tracking efficiency in jets to see if this uncertainty could be reduced.

Both the BEMC energy resolution and the TPC tracking efficiency contribute to the uncertainty of jet energy scale. Since BEMC resolution improves as deposited energy increases, at high p_T, TPC tracking efficiency represents the largest uncertainty on jet energy scale. Reducing the systematic uncertainty of TPC tracking efficiency will provide a more precise energy measurement of high p_T jets.

B. Simulation tools in STAR

1. Pythia, gSTAR and Monte-Carlo simulation data production

After RHIC Run 6 finished in June, 2006, a Monte-Carlo simulation data production [19] was made for the Run 6 detector configuration of STAR. This dataset was produced utilizing the Pythia [20] high energy collision event generator plus the gSTAR simulation kit [21], which is based on GEANT3 [22] and implements the STAR detector geometry. This dataset was used in this analysis as a counterpart of

the dataset of Run 6 longitudinal data.

In addition, the gSTAR simulation kit, which can simulate the energy deposition of primary particles in each part of the STAR detector, was used in this analysis to generate TPC responses for the single track event used in embedding.

2. The TPC response simulator

Besides the stand alone simulation tools, part of the STAR event reconstruction code also serves as part of the simulation process. Since gSTAR only simulates energy deposits of primary particles, the TPC response simulator, as part of the standard STAR code library, is used to carry on the simulation process and generate TPC ADC values from the energy deposition.

CHAPTER III

TPC TRACKING EFFICIENCY MEASUREMENT AND SYSTEMATIC UNCERTAINTY ESTIMATION

As charged primary particles travel through the TPC, they interact with the chamber gas and release secondary electrons along their trajectories. Once these secondary electrons drift through the chamber gas and reach the TPC anode planes, their position information in these planes is granularized[1] with respect to the positions of read out pads. These granularized clusters, together with their position information along the drift direction calculated from drift time, are referred as "TPC Hits" [12]. After that, tracks are reconstructed from the detected TPC hits by a software track fitting algorithm [23]. Then various cuts are applied to these reconstructed tracks in jet reconstruction and further analysis.

However, some tracks from a certain proton-proton collision event may not be reconstructed from the TPC hits, or may be excluded as "junk" tracks in the calculation of reconstructed jet energy. Listed here are some of the reasons that could cause the inefficiencies in TPC tracking [12, 13]:

1. Limited acceptance of the detector. There exist inactive areas of read out pads on the anode plane such as spaces used for mounting and fiducial cuts on the edge of TPC sectors. A track that intersects these areas may have too few detected TPC hits to be successfully reconstructed. For high momentum tracks, this will result in an inefficiency of $\sim$6%.

2. Limited resolution of the positions of TPC hits. Diffusion of the secondary elec-

[1] Along the radial direction of the TPC anode plane; though, along the azimuthal direction, positions of TPC hits are read out with multiple adjacent pads, thus acquiring a much better position resolution than the width of the pads.

trons when they drift toward the anode wire planes, finite size of read out pads, finite resolution of drift time measurement and noises of front end electronics all lower the resolution of the positions of detected TPC hits. For example, the dimension of the read out pads along the pad row is 6.20 mm for the outer sub-sectors; though the position resolution is improved by reading out three adjacent pads at the same time, the finite size of read out pads still introduces $\sim$2 mm uncertainty on the position of TPC hits along pad rows for tracks with transverse momentum of 200 MeV.

The smeared positions of TPC hits deteriorate the quality of reconstructed tracks. The measured properties of a reconstructed track, such as the distance of closest approach[2], will have a deviation from the true value introduced by the uncertainty of TPC hits and fail to pass certain cuts in an analysis. Since these cuts are only supposed to remove tracks that should not be counted in a reconstructed event, for example, pile-up tracks and double counted split tracks, the accidental removal of primary tracks with poor reconstruction quality should be considered as an inefficiency of TPC tracking.

3. Merging of TPC hits. Diffusion of the drifting secondary electrons of a track causes another problem. Though reading out adjacent pads improves the position resolution of TPC hits, this strategy does not work in the situation when secondary electron clusters from two different tracks overlap each other—the distance between two clusters needs to be much larger than the resolution mentioned above to be successfully separated into two TPC hits. Merging of TPC hits reduces the available TPC hits for track reconstruction and causes merging

[2]The distance from event vertex to the reconstructed track. Will be discussed in Ch. IV.

of two adjacent tracks.

These various reasons that could cause the failure of reconstructing a track make an analytical tracking efficiency calculation infeasible. Furthermore, the high luminosity condition and the relatively slow TPC response makes this analysis even more difficult [3]. Consequently, the technique of embedding, which calculates the tracking efficiency by Monte-Carlo simulation while utilizing real data, is used in this analysis and discussed in this chapter.

A. The Embedding Technique

A detector simulator builds models of physical processes to produce reasonable and realistic detector responses. The simulation of TPC responses is divided into two steps. First, the gSTAR/GEANT3 simulation kit produces secondary particles that were released by the simulated primary tracks and calculates energy depositions as charged tracks pass through matter. Then, codes in the standard STAR analysis chain, the StTrsMaker, simulate other processes in the TPC that could not be handled by GEANT and produce raw Analog/Digital Converter (ADC) values that correspond to what was recorded in real data[4]. The combined simulation of gSTAR plus StTrsMaker takes care of many processes that could affect the TPC tracking, such as multiple collisions between primary particles and detector material, diffusion of secondary electrons during drifting, gain fluctuations of the electron avalanche near MWPC anode wires, and responses of the TPC read out pads and signal shapers [13].

---

[3]The influence of the beam luminosity on TPC tracking will be discussed later in this chapter.

[4]In fact, only a small fraction of runs recorded the TPC ADC values to the tape; most of the time, only real-time reconstructed TPC hits were saved while ADC information was dropped.

Therefore, the physical processes that could affect TPC hit reconstruction are considered in the simulation and it is applicable to the task of the TPC tracking efficiency calculation.

One possible method of estimating the TPC tracking efficiency is comparing the tracks in the simulation data before and after the simulation and event reconstruction. For each event, the tracking efficiency is calculated as the number of simulated tracks that could be associated to a reconstructed track, divided by the total number of simulated tracks. The standard Run 6 Monte-Carlo simulation data (pure simulation) itself can serve such kind of analysis and produce valid calculations of the tracking efficiency, as long as pure simulation provides a good approximation of true data.

Unfortunately, however, in practice the TPC works in a much more complex and harsher environment than the one in pure simulation. Given the facts that the distance between the TPC central membrane and each anode wire plane is 210 cm, and that the average gas drift velocity is 5.45 cm μs^{-1}, the read-out time window of the TPC is $\sim$40 µs, which is much longer than the RHIC beam bunch crossing interval of $\sim$106 ns. At a BBC coincidence rate of 400 kHz, there are approximately 32 collisions that lie within this time window before and after the triggering event, which are referred as "pile-up" events. Segments of tracks from pile-up events will be read out together with tracks from the triggering event, disturbing the TPC tracking. Since the TPC drift in the beam direction preserves the relative transverse location of hit positions along a track, out-of-time tracks still look like tracks, but maybe with hits lost from one end or another. On average in Run 6, there were 1.4×10^2 pile-up tracks [24] with transverse momentum greater than 0.2 GeV in each event. Figure 6 illustrates two pile-up events before and after a triggering event.

Besides the pile-up tracks, another factor that pure simulation does not consider is the background from beam-gas interactions, which could also affect the TPC

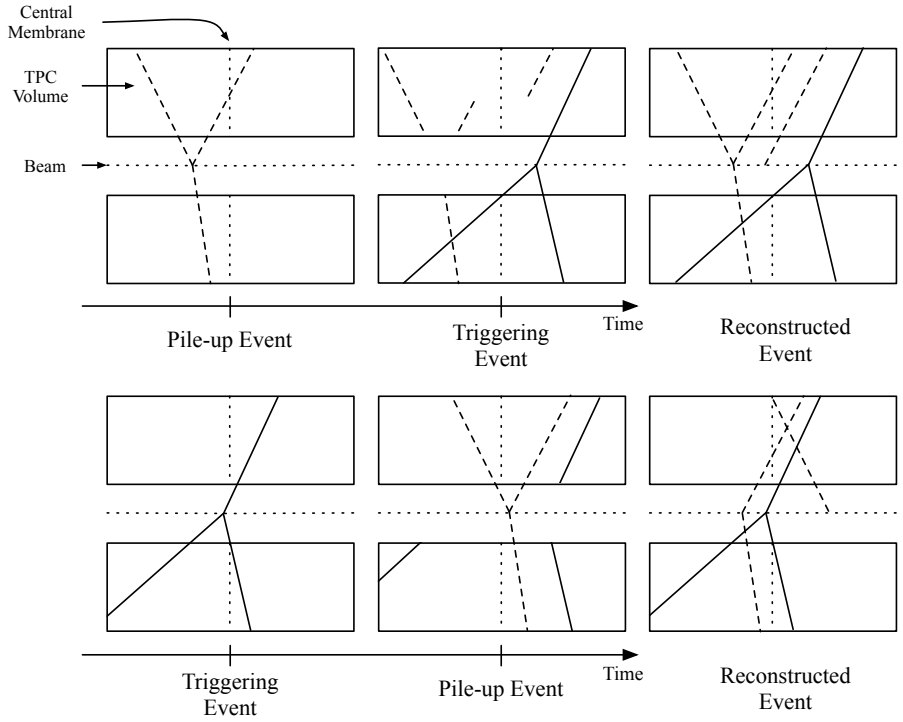


Fig. 6. Tracks from pile-up events before and after the triggering event.

tracking efficiency. In order to address these issues, the *embedding* technique was used in this analysis, in which detector responses from a (single track) pure simulation event were mixed with those from a real data event. Since only one simulated track is used in each event, the mixed event looks almost identical to the events in real data in terms of track transverse momentum distribution, reconstructed jet energy distribution, shape of reconstructed jet, and etc.

1. Event mixing between single track simulation and real data

For each event from real data used for embedding, a single track event was simulated to obtain TPC responses. In this analysis, only charged pions (π^+ and π^-) are considered in the single track simulation, since most of the charged particles in the

reconstructed jets are charged pions.

The single track simulation event is embedded into an event from real data at the TPC ADC level. After a simulated track is handled by gSTAR and the TPC response simulator, the obtained TPC ADC values are added to the ones from a real data event. The mixed event then is fed to the event reconstruction code as a whole. The TPC hits, the tracks and the jets are then reconstructed from the mixed TPC ADC values step by step. The effects of the high luminosity environment on the TPC tracking is reflected on the distorted TPC ADC values associated with the TPC hits along the simulated track.

One major limitation on the statistical errors of this analysis is the fact that only a small portion of the real data have the TPC ADC information[4].

2. Estimating Systematic Uncertainty of the Tracking Efficiency

For a measurement of a certain value, systematic uncertainty is commonly defined as the difference between the true value and the average of measured values when the number of measurements approach infinity, which is the bias introduced by the inaccurate understanding of the measurement. In this analysis of TPC tracking efficiency, this bias could be estimated by introducing the embedding into pure simulation data (Monte-Carlo embedding), in addition to the embedding into real data (real data embedding). The simulation is essentially a calculation based on our understanding of the measuring equipment, thus the difference between the calculations of MC embedding and real data embedding reflects the bias on the understanding of how the environment of the TPC volume affects the tracking efficiency.

3. Event mixing between single track simulation and Monte-Carlo data

In the embedding into Monte-Carlo data, the mixing is performed right after the secondary electrons along the primary tracks are simulated by gSTAR, and before the event is fed to the TPC response simulator. Despite this difference of when the events are mixed, other procedures in the real data embedding and Monte-Carlo data embedding are kept as same as possible in this analysis to prevent introducing any other differences between these two methods instead of those that are intrinsic.

B. Issues in the Embedding Procedure

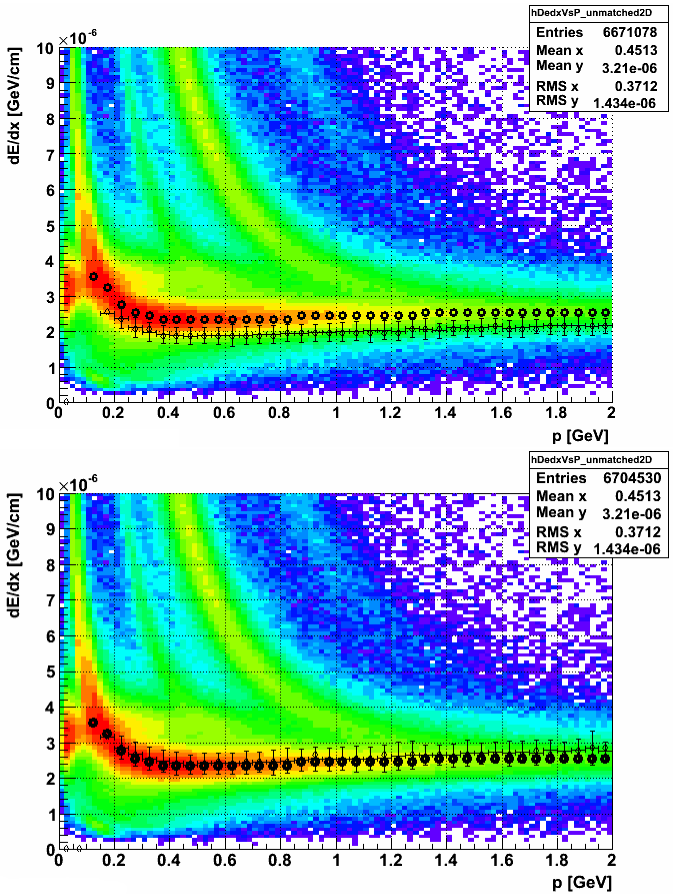
1. Ionization energy loss values and the "fudge factor" in single track simulation.

One issue in the simulation is that the value of ionization energy loss per unit distance traveled by charged particles, $d\mathrm{E}/d\mathrm{x}$, is not consistant with real data. Figure 7 shows the differences in $d\mathrm{E}/d\mathrm{x}$ vs. track momentum. This inconsistency could be introduced by various reasons, such as inaccurate $d\mathrm{E}/d\mathrm{x}$ values, or inaccurate conversion from $d\mathrm{E}/d\mathrm{x}$ to ADC values.

To fix this issue, energy deposit values from GEANT are multiplied by a "fudge factor" to match the values in real data. In this analysis, the fudge factor was determined by matching the pion peaks from simulation and real data in the $d\mathrm{E}/d\mathrm{x}$ vs. track momentum plot, and the value of 1.26 was applied.

2. TPC gas gain fluctuations.

Another issue related to fudge factor in real data embedding is the fluctuations in TPC gas gain. The TPC pressure is regulated to be 2 mbar above atmospheric pressure [12], and fluctuates as weather conditions change; so does the TPC gas gain which is a function of gas pressure. In the experiment, the gas gain is monitored to



Circle: poin peak of real data;
Diamond: pion peak of simultion; Bars: pion peak width of simulation.

Fig. 7. Differences in energy loss values between simulation and real data, before and after fudge factor applied.

calibrate TPC data to get constant dE/dx value; however, tracking efficiency is also affected but cannot be reflected in the calibration. As shown in Figure 8 [25], gas gain variation over the running period of Run 6 was within ±6%.

In the simulation, the effect of gas gain fluctuation on the tracking efficiency can be addressed by varying the fudge factor the same amount. Varying the fudge factor will change the ADC values from the simulation, which is equivalent to the effects of gas gain variation.

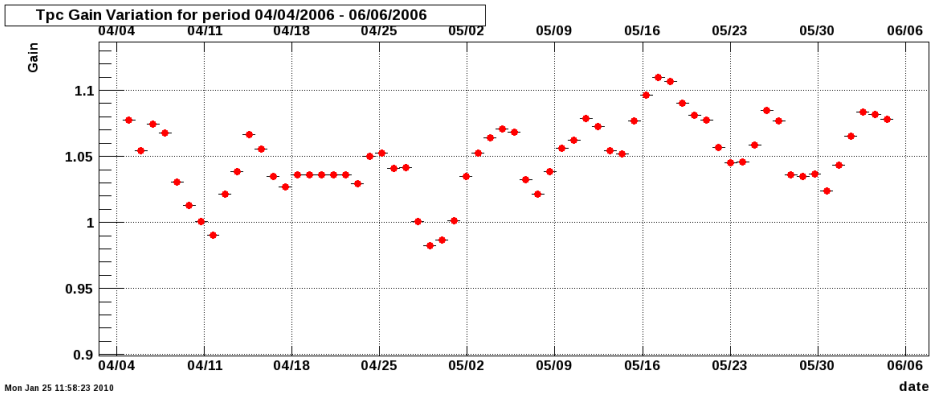


Fig. 8. The TPC gas gain variation during the running period of Run 6.

3. Beam luminosity dependence.

One major difference between simulation and real data is the presence of pile-up events. As discussed in the previous section, the average number of pile-up tracks in a triggering event is proportional to the BBC coincidence rate. It is expected that the difference between simulation and real data drops at lower BBC rates, and the effects of pile-up tracks could be much reduced by fitting the tracking efficiency as a function of BBC rate and using the extrapolated value at zero luminosity to compare with simulation.

However, in this analysis, this approach did not work due to limited event-by-event variation in BBC rate. Figure 9 demonstrates the BBC rate distribution of the events. An attempt to fit tracking efficiency against BBC rates showed that the correlation between tracking efficiency and BBC rate was buried in statistical fluctuations.

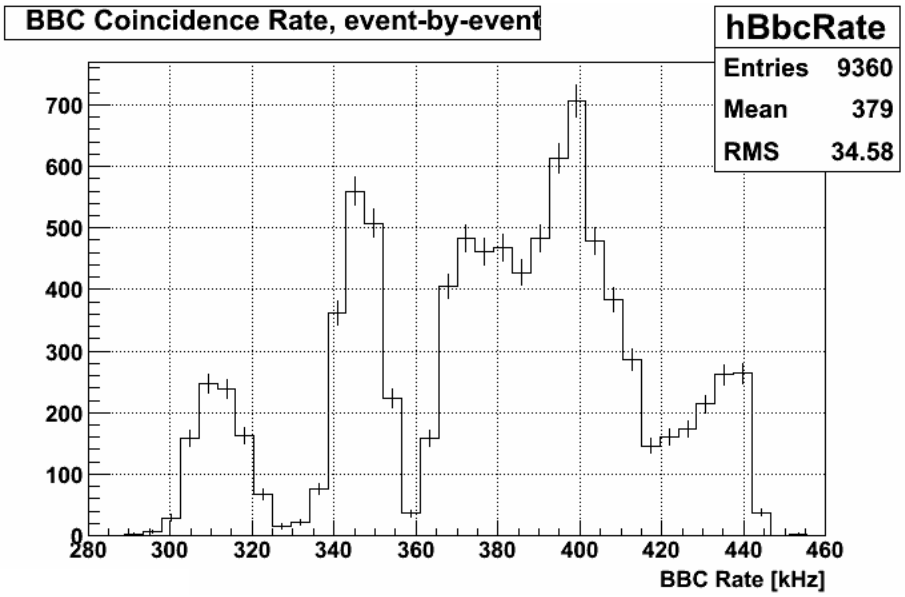


Fig. 9. BBC coincidence rate distribution. (Events in real data used for embedding.)

CHAPTER IV

IMPLEMENTATIONS OF THE EMBEDDING PROCEDURE

This chapter explains how the embedding procedure is implemented in this analysis. In this chapter certain terms are defined as follows:

Embedding Event/Jet An event/jet into which the simulated track is embedded.

Embedded Track A single track that is simulated, mixed with embedded event, and later reconstructed to evaluate the tracking efficiency.

Sampled Event/Jet An event/jet from which the kinematics of an embedding track are sampled.

The implementation is done as far as possible in the same way for both real data and MC embedding, while the largest difference is found in how the embedded tracks and embedding events are mixed, which will be discussed in Sec. C.

The kinematics of the embedding track are sampled randomly from Run 6 200 GeV p+p longitudinal data, for both real data and MC embedding, based on the kinematics of the embedded jet, which are obtained from reconstructed embedding events. Applying these kinematics, a single track is simulated by GSTAR and its TPC responses are mixed with responses from the embedding event, which are then reconstructed as a whole event. Associations are then attempted between each primary track in the reconstructed event and the embedded track. If no primary track can be associated to the embedded track, then the embedded track will be considered missing due the inefficiency of TPC tracking. Fig. 10 illustrates the procedures of both the real data and MC embedding. Details of each step will be discussed in this chapter.

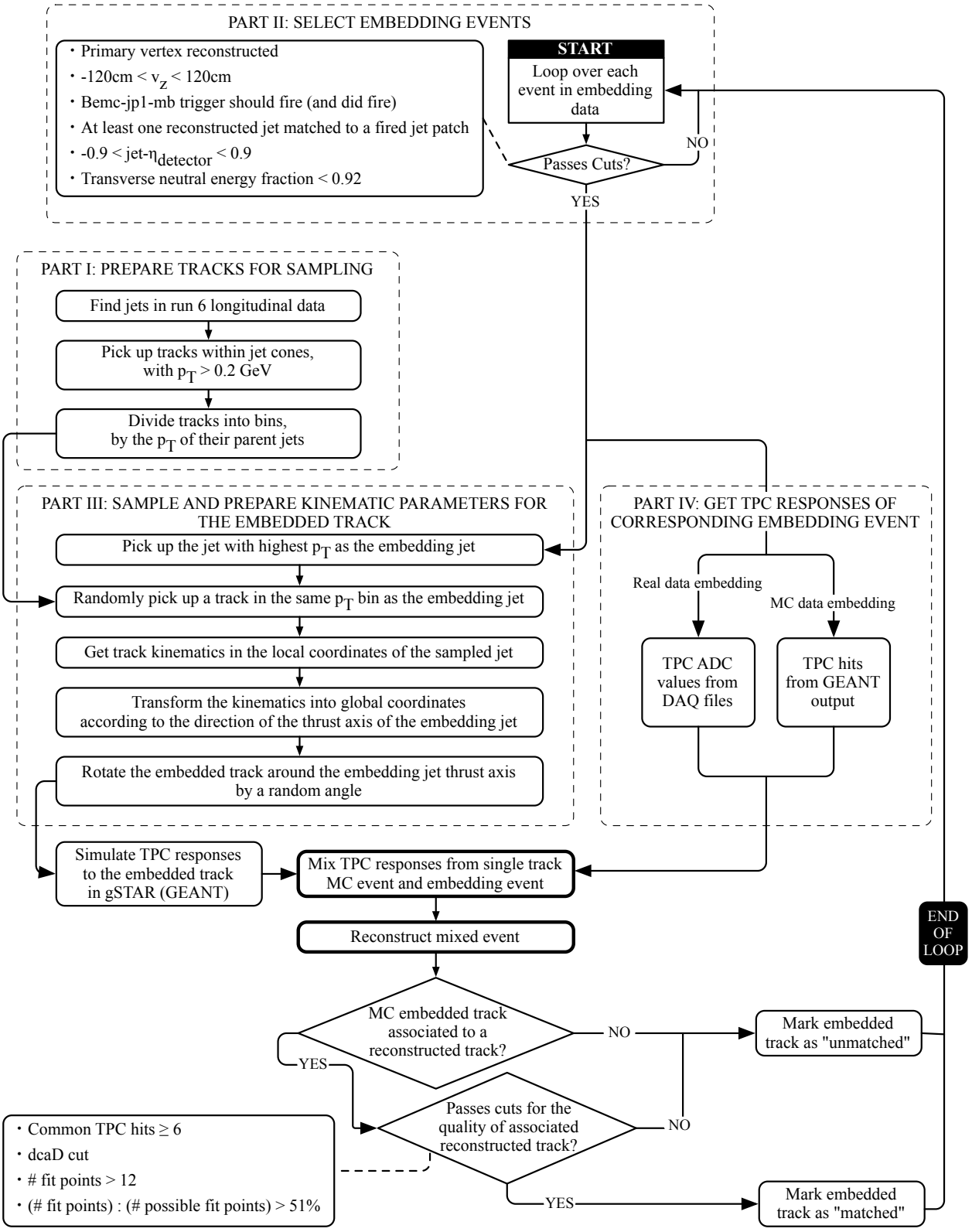


Fig. 10. Procedures for real data and MC embedding

A. Datasets Used for Analysis

Data sets both before and after reconstruction are used in this analysis. In part III of Fig. 10, kinematics of embedding jets, which are used in sampling embedded track kinematics, are extracted from reconstructed events, whereas the TPC responses, which are used in event mixing, are extracted from corresponding embedding events in datasets before reconstruction. Run 6 200 GeV p+p longitudinal data after reconstruction are also used in sampling kinematics of embedded tracks, in both real data and MC embedding.

1. Cuts on the Embedding Events

The analysis code loops over each event in the available reconstructed data sets and selects events that pass the following cuts for embedding. These cuts separate out events that contain at least one reconstructed jet located in the region with better TPC performance. Furthermore, similar cuts are also commonly found in other jet analyses [15, 16, 26] which makes this analysis more consistent with others.

(1) A primary vertex reconstructed. In this analysis, if a track cannot be associated to the primary vertex after reconstruction, it will be considered as an inefficiency of the TPC tracking, even though there might be a non-primary track reconstructed. Therefore only events with a primary vertex are considered.

(2) -120 cm $< v_z <$ 120 cm, where v_z is the primary vertex position with respect to the detector center along the beam line. This cut eliminates events in which the collision happened near the boundary of TPC.

(3) BEMC-JP1-MB trigger "should fire" in both real data and MC embedding. BEMC-JP1-MB trigger "did fire" in real data embedding. BEMC-JP1-MB re-

quires a minimum energy deposit of 8.3 GeV in a single jet patch, which is defined as a BEMC region of size 1×1 in η and ϕ. In the real data embedding, this trigger contributes most of the jet events recorded in the Run 6 200 GeV p+p longitudinal data and the "should fire" requirement acts as a software trigger. On the other hand, in MC embedding, this same trigger filters the events and makes sure that the behaviors of this trigger are also reflected in the MC data.

(4) At least one reconstructed jet matched to a fired jet patch, by requiring that the difference between the center of the jet patch and the jet thrust axis is smaller than 0.6 in η and ϕ.

(5) $-0.9 <$ jet-$\eta_{\text{detector}} < 0.9$. Given the fact that when tracks are required to have at least 12 fit points, the TPC only has acceptable tracking efficiency approximately between $-1.3 < \eta_{\text{detector}} < 1.3$, this cut ensures that most tracks in a jet reconstructed with a cone radius of 0.7 fall in the TPC coverage.

(6) Transverse neutral energy fraction of the jet < 0.92. This value is the ratio of the transverse energy deposited in EMC to the total transverse energy of the jet. This cut ensures that TPC tracks contribute a significant amount of transverse energy to the reconstructed jet, which suppresses fake jets from non-collision backgrounds.

2. Statistics of Embedding Events in Real Data Embedding

In real data embedding this analysis mainly focuses on the Run 6 200 GeV p+p longitudinal data. In part IV of the embedding procedure illustrated in Fig. 10, raw information of TPC responses in the form of TPC ADC values are required to mix the embedding event with the embedded track. In the recorded data, TPC ADC information can be found in DAQ files, which were obtained from the Data

AcQuisition system during the run. They were recorded for each pad as a time series of size 512 for each of the 5692 pads on the TPC endcaps [12].

However, due to limited tape writing rate, most events recorded during the run do not contain TPC ADC information. Instead, only online reconstructed clusters and tracks were written to the tape. Only a subset of DAQ files from 15 runs in the Run 6 200 GeV p+p longitudinal data contain TPC ADC information, which totals ∼65 k. After applying the above cuts, there remains ∼9 k events available for embedding. The limited number of events available for real data embedding is the major contribution to the statistical uncertainty of this analysis.

3. Reweighting of the Embedding Events in MC Embedding

Data sets from the P07ic pp 200 GeV Monte Carlo production series are used in the MC embedding. These data sets were produced with PYTHIA v6.410 and year 2006c geometry[1] of the detector and were divided into different partonic p_T bins from 3 GeV to 65 GeV.

One difficulty in the Monte Carlo simulation arises from the steeply falling spectrum of the partonic p_T distribution. The partonic p_T spectrum of participating partons in the collisions decays exponentially and could change several orders of magnitude in the range of interest. On one hand, many of the interesting events have larger partonic p_T, such as events with larger jet p_T and events that pass jet patch triggers; thus good statistics are desired on these events. On the other hand, events with smaller partonic p_T cannot be ignored because of the limited detector energy resolution and the steeply falling partonic p_T spectrum. For example, even with a minimum requirement on the reconstructed jet p_T, events from a partonic p_T bin that

[1] http://drupal.star.bnl.gov/STAR/comp/prod/monte-carlo-production-datasets

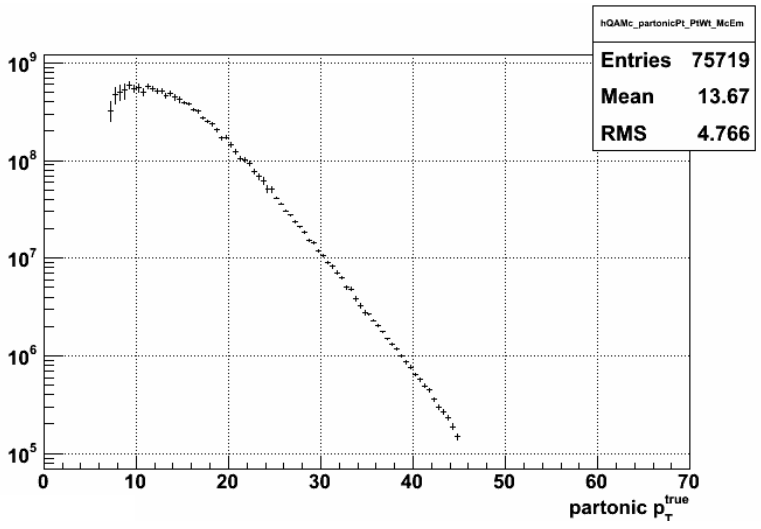


Fig. 11. Partonic p_T distributions of embedding events in MC embedding, after applying the cuts described in Sec. 1

is smaller than that threshold could still contribute significantly in the result. Considering the steeply falling spectrum of partonic p_T, if enough events in large partonic p_T bins are simulated to get good statistics, the number of events that need to be simulated in the small partonic p_T bins will become overwhelming and not feasible because of limited CPU time and storage space. To partially solve this conflict, a similar number of events in each partonic p_T bin are simulated and then get weighted during the analysis to recover the exponential decaying spectrum, which is referred to as the reweighting procedure.

In the MC embedding a portion of events from the available data sets in the partonic p_T bins from 7 GeV to 65 GeV was analyzed. For each partonic p_T bin, the number of events, number of events that pass the cuts in Sec. 1 and the corresponding reweighting factors are listed in Table I. Fig. 11 illustrates the distribution of partonic p_T after reweighting.

partonic p_T [GeV]	# of events	passed cuts	reweighting factor (ω)
7 – 9	420 k	0.6 k	4.08×10^6
9 – 11	412 k	2.9 k	1.02×10^6
11 – 15	420 k	13 k	4.03×10^5
15 – 25	397 k	58 k	7.25×10^4
25 – 35	400 k	171 k	14.4×10^3
35 – 45	120 k	48 k	3.73×10^2
45 – 55	120 k	63 k	2.04×10^1
55 – 65	118 k	58 k	1

Table I. Pure Simulation Dataset.

4. Statistical Uncertainty Calculations with Reweighting

The efficiency of TPC tracking is defined as the ratio of number of reconstructed tracks to the number of embedded tracks,

$$\epsilon = \frac{n_{\text{reconstructed}}}{n_{\text{embedded}}} \equiv \frac{n_+}{n}. \tag{4.1}$$

Assume n is fixed and n_+ follows binomial distribution [27]

$$P(n_+; n) = \binom{n}{n_+} p^{n_+} (1-p)^{(n-n_+)} \tag{4.2}$$

where p is the true efficiency, with standard deviation

$$\sigma_{n_+} = \sqrt{np(1-p)}, \tag{4.3}$$

then with ϵ as an estimation of true efficiency p, the standard deviation of ϵ could be calculated as

$$\sigma_\epsilon = \frac{\sigma_{n_+}}{n} = \sqrt{\frac{\epsilon(1-\epsilon)}{n}}. \tag{4.4}$$

Notice that this formula only works fine when ϵ is a good estimation of p, which requires that n, ϵn, and $(1-\epsilon)n$ are all large enough.

When reweighting is involved, the definition for the efficiency becomes

$$\epsilon = \frac{\sum_i \omega_i n_{i+}}{\sum_i \omega_i n_{i+} + \sum_i \omega_i n_{i-}}, \tag{4.5}$$

where i is the index for different partonic p_T bins and $\sum$ sum over all considered partonic p_T bins. ω_i's are the factors for reweighting and n_{i-} is defined as

$$n_{i-} \equiv n - n_{i+}. \tag{4.6}$$

Assume n_{i+}, n_{i-} are independent and are large enough to be considered as Poisson distributed, with standard error propagation equation

$$\sigma_\epsilon = \sqrt{\sum_i (\frac{\partial \epsilon}{\partial n_{i+}})^2 \sigma^2_{n_{i+}} + \sum_i (\frac{\partial \epsilon}{\partial n_{i-}})^2 \sigma^2_{n_{i-}}}, \tag{4.7}$$

which gives

$$\sigma_\epsilon = \frac{\sqrt{(\sum_i \omega_i^2 n_{i+})(\sum_i \omega_i n_{i-})^2 + (\sum_i \omega_i^2 n_{i-})(\sum_i \omega_i n_{i+})^2}}{(\sum_i \omega_i n_{i+} + \sum_i \omega_i n_{i-})^2} \tag{4.8}$$

B. Kinematics of the Embedded Tracks

1. Sampling Track Kinematics by Jet p_T of Embedding jets

In both real data and MC embedding, embedded track kinematics are sampled from events in Run 6 200 GeV p+p longitudinal data. These events are picked from the same runs as the embedding events, but without TPC ADC information and not used

as an embedding event. In real data embedding, tracks are sampled from the events in the same run as the embedding event, while in MC embedding, tracks are sampled from all available events.

All the tracks found within jet cones of the reconstructed events were sampled and divided into different bins by the transverse momentum of the sampled jet. Table II show how many tracks are available for sampling in each jet p_T bin. When an embedding event is selected, a track is randomly selected from the bin with the same p_T of the embedding jet. In this way, the sampled kinematics reflect track distributions related to jet p_T, such as track p_T and distance from the jet thrust axis.

jet p_T range [GeV]		# of tracks	jet p_T range [GeV]		# of tracks
7.56	9.30	33914	21.3	26.2	67279
9.30	11.4	107921	26.2	32.2	23763
11.4	14.1	186008	32.2	39.6	6343
14.1	17.3	194216	39.6	48.7	1076
17.3	21.3	137005			

Table II. Number of tracks in each jet p_T bin available for sampling embedded track kinematics.

2. Local Coordinate System for Jets

A local jet coordinate system is defined to describe track kinematics with respect to the jet thrust axis, which is illustrated in Fig. 12:

$\hat{\mathbf{L}}$ Parallel to jet thrust axis.

$\hat{\mathbf{N}} \equiv \frac{\hat{\mathbf{z}}\times\hat{\mathbf{L}}}{|\hat{\mathbf{z}}\times\hat{\mathbf{L}}|}$ Perpendicular to reaction plane.

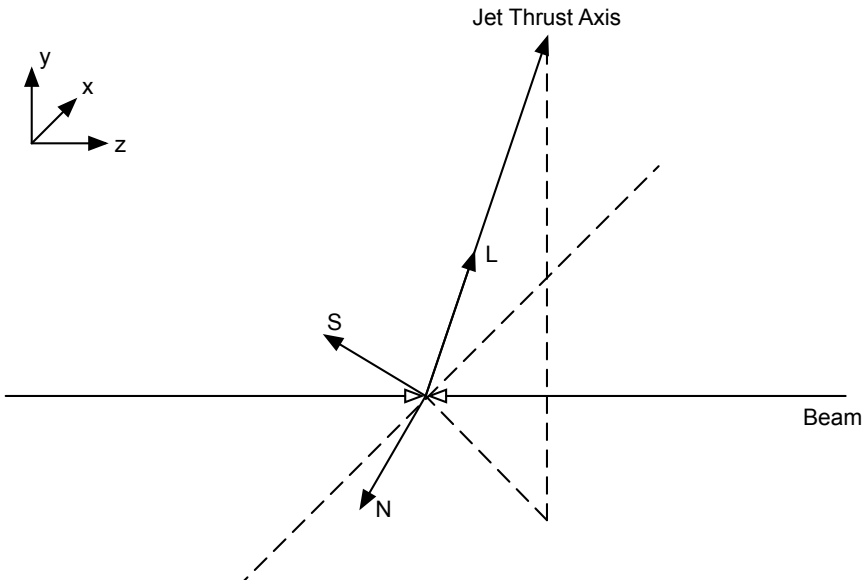


Fig. 12. Definition of Local Jet Coordinate System

$\hat{\mathbf{S}} \equiv \hat{\mathbf{N}} \times \hat{\mathbf{L}}$ Sideways unit vector.

The kinematics of embedded tracks are sampled in the local jet coordinate system as

$$p_S = \vec{\mathbf{p}}_\mathbf{0} \cdot \hat{\mathbf{S}}_\mathbf{0},\ p_N = \vec{\mathbf{p}}_\mathbf{0} \cdot \hat{\mathbf{N}}_\mathbf{0},\ p_L = \vec{\mathbf{p}}_\mathbf{0} \cdot \hat{\mathbf{L}}_\mathbf{0}$$

where $\vec{\mathbf{p}}_\mathbf{0}$ is the track momentum with respect to the global coordinates in the sampled event and $\hat{\mathbf{S}}_\mathbf{0}$, $\hat{\mathbf{N}}_\mathbf{0}$, and $\hat{\mathbf{L}}_\mathbf{0}$ are local coordinate axes of the sampled jet. Then the kinematics are transformed to the embedding event by

$$\vec{\mathbf{p}}' = p_S \cdot \hat{\mathbf{S}}' + p_N \cdot \hat{\mathbf{N}}' + p_L \cdot \hat{\mathbf{L}}'$$

where $\hat{\mathbf{S}}'$, $\hat{\mathbf{N}}'$, and $\hat{\mathbf{L}}'$ are local coordinate axes of the jet in the embedding event. The last step in determining the kinematics is to rotate the embedded track around the embedding jet axis for a random angle between 0 to 2π. This rotation is introduced to eliminate the uneven track distribution in the azimuthal direction crossing different

jet patches. As illustrated in Figure 13, in the azimuthal direction, a jet patch spans $\frac{\pi}{3}$ and a TPC sector spans $\frac{\pi}{6}$, with the center of each jet patch overlapping centers of every other TPC sector, resulting a $\frac{\pi}{12}$ distance between jet patch centers and TPC sector boundaries. Since the jet thrust axis distribution for triggered jets is highly weighed at the center of each jet patch, less tracks are sampled at TPC sector boundaries where the tracking efficiency is most affected by the inactive areas of the TPC. The randomized rotation fills the gaps near TPC sector boundaries of the sampled track distribution and ensures that the analysis reflects the overall TPC tracking efficiencies, instead of those within TPC sectors.

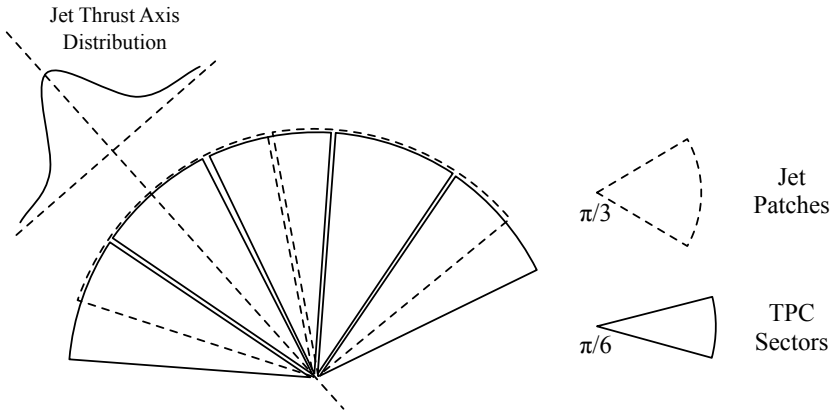


Fig. 13. Jet Patches and Corresponding TPC Sectors on the TPC Endcap.

C. Single Track Simulation, Event Mixing and Event Reconstruction

The details of single track simulation and event mixing in real data and MC embedding is explained in Ch. III Sec. A. In real data embedding, the simulated TPC ADC values of the single track event were added to those in the real data event, and then the event was constructed as if it comes from real data. In contrast, in MC embedding, the TPC responses from the single track simulation were mixed with the

event from the Monte-Carlo data as TPC hits. Then the mixed TPC hits were fed to the TPC response simulator to get the TPC ADC values for further reconstruction.

D. Association of Tracks from Reconstructed Event and MC Event

After the reconstruction of the mixed event, for each reconstructed track, an attempt is made to associate it to the embedded track in the Monte Carlo event through the StAssociationMaker[2]. Tracks are associated in the StAssociationMaker if they have a minimum number of common TPC hits. In this analysis, this minimum is set to 6 and TPC hits in the reconstructed event are associated with hits in the embedded track if the distance of the hits in X and Y is smaller than 5 mm and in Z is smaller than 2 mm.

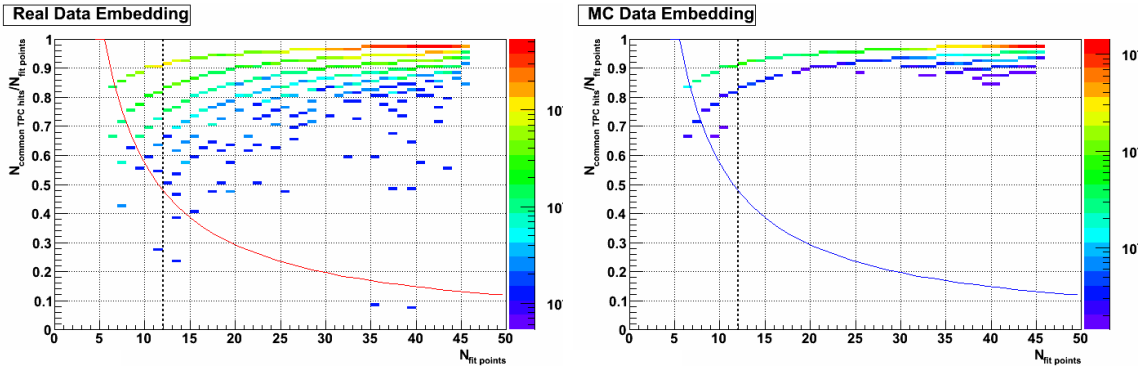


Fig. 14. Effects of the Cut on the Number of Common TPC Hits.

Because of the inefficiency of the association process, there might be reconstructed embedding tracks not able to be associated to the embedded tracks and thus identified as not reconstructed. However, as illustrated in Fig. 14, for both real data and MC embedding, only a few associated tracks were dropped[3]. For this reason,

[2]Reference:STAR Offline Library Long Writeup : StAssociationMaker User Guide and Reference Manual

[3]When combined with the number of fit points > 12 cut in next section

it is assumed that the track association inefficiency is negligible in this analysis.

E. Applying Cuts on the Associated Reconstructed Tracks

Before an associated reconstructed track is counted as a successfully reproduced track from the TPC, several more cuts are applied. These cuts are standard in jet finders [15, 16, 26] and are implemented to eliminate pile-up tracks and split tracks and improve track quality. However, these cuts are imperfect and also cut off tracks that are reconstructed from real tracks in jets and reduce the actual TPC tracking efficiency. A summary of reduced tracking efficiency in real data and MC embedding is shown in Table III.

	Real Data Embedding	MC Embedding	Difference
dcaD cut	-3.8%	-2.6%	1.2%
number of fit points > 12	-3.0%	-1.8%	1.2%
$\frac{\text{number of fit points}}{\text{number of possible fit points}} > 51\%$	-0.5%	-0.2%	0.3%
Total difference in reduction			2.7%

The cuts are applied sequentially, in the order given.

Table III. Summary of reduced tracking efficiency due to cuts on reconstructed tracks.

1. Distance of Closest Approach

The Distance of Closest Approach (DCA) cut limits the distance between a track and the primary vertex of an reconstructed event, and is implemented to remove pile-up tracks which are irrelevant to the primary vertex. dcaD is the transverse component of

DCA. The standard cut used in Run 6 jet analysis is dependent on track momentum

$$dcaD <= \begin{cases} 2 \text{ cm} & \text{if } p_T < 0.5 \text{ GeV} \\ (3 - 2 \times p_T/\text{GeV})(\text{cm}) & \text{if } 0.5 \text{ GeV} < p_T < 1 \text{ GeV} \\ 1 \text{ cm} & \text{if } p_T > 1 \text{ GeV}. \end{cases}$$

Fig. 15 illustrates the track momentum dependent dcaD cut.

The reduction of the tracking efficiency by this cut in real data embedding is 1.2% more than that in MC embedding, which indicates that the associated reconstructed tracks in MC embedding have better track quality.

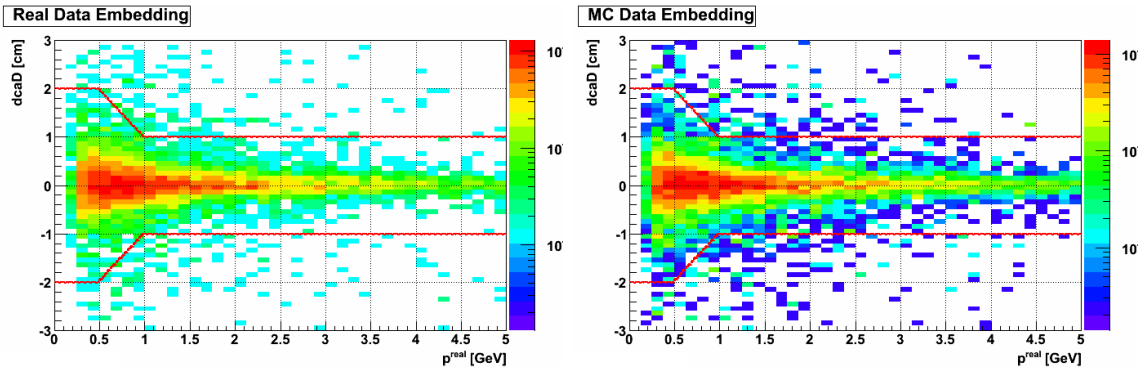


Fig. 15. dcaD cut vs. track momentum.

2. Cuts on the Number of Fit Points

These cuts require that a track should at least have 12 fit points and that the number of fit points divided by the number of "possible fit points"[4] should be greater than fifty one percent. Both of these cuts remove tracks that are likely to be reconstructed with poor resolution. In addition, these cuts reduce the number of pile-up tracks that typically are reconstructed using only a subset of the original hits, as well as double

[4] "Possible fit points" are defined as the maximum possible number of TPC read out pads that a track could pass over, which are calculated based on track kinematics and the active electronics channels.

counted split tracks, since the two tracks in a split track pair cannot both own more than half of the possible fit points.

Fig. 16 illustrates the effects of these two cuts. Comparing to MC embedding, tracking efficiency reduced about 1.5% more in real data embedding because of these two cuts, indicating that reconstructed tracks in real data embedding have lower quality and a larger portion of them are accidentally rejected by the cuts as pile-up tracks and split tracks.

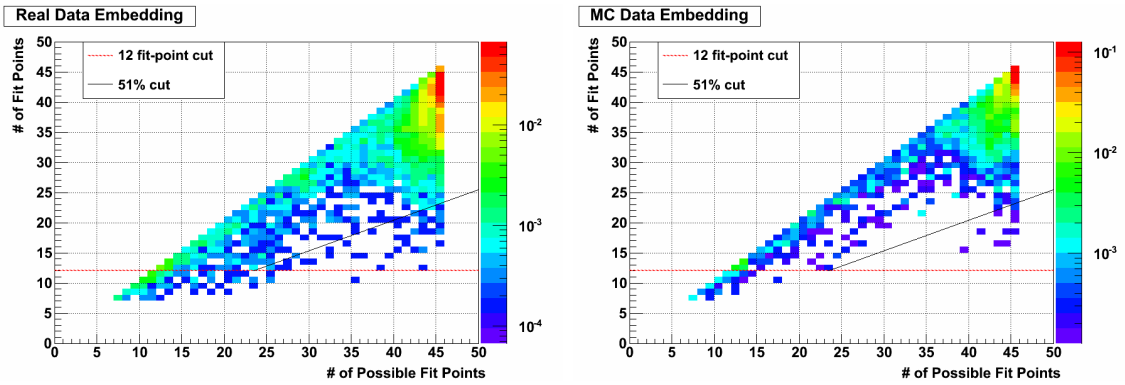


Fig. 16. Cuts on the Number of Fit Points

CHAPTER V

ANALYSIS RESULTS AND CONCLUSION

A. Comparison of Calculated TPC Tracking Efficiencies between Real Data Embedding and MC Embedding

This section discusses the calculation results of TPC tracking efficiencies in both real data embedding and MC embedding, based on the datasets and analysis procedures discussed in previous chapters, for the STAR 200 GeV proton-proton longitudinal data during Run 6. In general, tracking efficiencies calculated in real data embedding are systematically lower than those calculated in MC embedding.

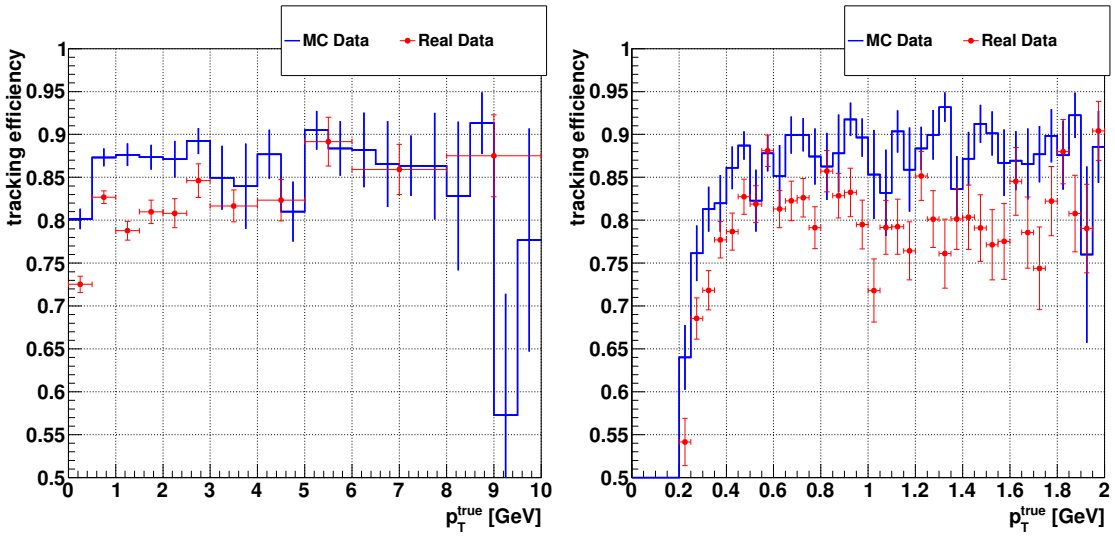


Fig. 17. Tracking efficiencies vs. transverse momentum (p_T) of embedded tracks

Figure 17 shows tracking efficiencies vs. transverse momentum of embedded tracks. The plot on the left is for tracks with p_T between 0.2 GeV and 10 GeV, while the right one is for low p_T tracks between 0.2 GeV and 2 GeV. The result indicates that the tracking efficiencies are relatively constant when track p_T is greater

than ∼1 GeV, with only fluctuations comparable to statistical errors, in both real data embedding and MC embedding. Tracking efficiencies start to fall drastically when p_T is lower than 0.5 GeV, as multiple scattering increases, tracks are bent by the magnetic field more significantly, and tracks become more unlikely to be reconstructed properly. Though the calculated efficiency in real data embedding is consistently lower than the one in MC data embedding, the shape of the efficiency curves matches well with each other.

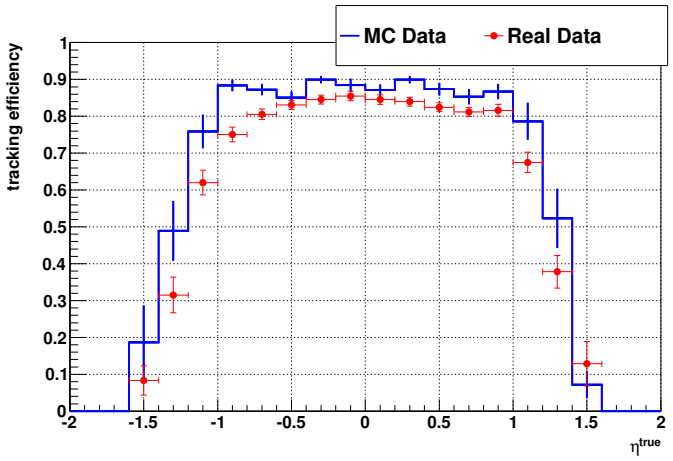


Fig. 18. Tracking efficiencies vs. pseudo-rapidity (η) of embedded tracks

Figure 18 shows the dependencies of tracking efficiencies on the track pseudo-rapidity. At mid-rapidity, tracking efficiencies are relatively independent on track pseudo-rapidity, for both efficiencies calculated from real data embedding and MC embedding. As track pseudo-rapidity reaches ±1, tracks start to leave the TPC volume through the TPC endcaps, resulting in shorter tracks that can be reconstructed. As a result, tracking efficiencies start to fall rapidly beyond ±1 in pseudo-rapidity, and the TPC loses its tracking ability when track pseudo-rapidity is larger than ∼ 1.5. Similar to the tracking efficiency dependence on transverse momentum, tracking efficiency calculated in real data is consistently lower than in MC embedding, although

the shapes match.

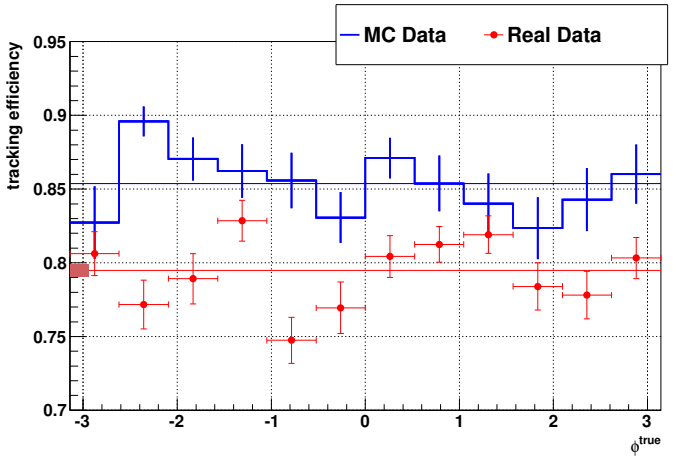


Fig. 19. Tracking efficiencies vs. azimuthal angle (ϕ) of embedded tracks

Figure 19 shows the tracking efficiencies vs. azimuthal angle of embedded tracks. The plot is binned in ϕ corresponding to TPC sector boundaries, and indicates inconsistencies of tracking efficiency across different TPC sectors that cannot be explained by statistical fluctuation. The solid lines indicate average tracking efficiencies calculated from dividing the number of total matched reconstructed tracks by the number of total embedded tracks. The average tracking efficiencies are shown in Table IV.

	Efficiency	Difference
Real Data Embedding	$79.5 \pm 0.4\%$	
MC Embedding	$85.4 \pm 0.5\%$	
		$5.9 \pm 0.6\%$

Table IV. Comparison of tracking efficiencies.

To see whether this difference between real data embedding and MC embedding could be reduced when only considering the p_T and η ranges where the TPC has

better tracking performances, two tighter cuts, $p_T > 0.5$ GeV and $-1 < \eta < 1$, are applied to the previous results. Figures 20, 21 and 22 demonstrate the tracking efficiencies when these additional cuts are applied.

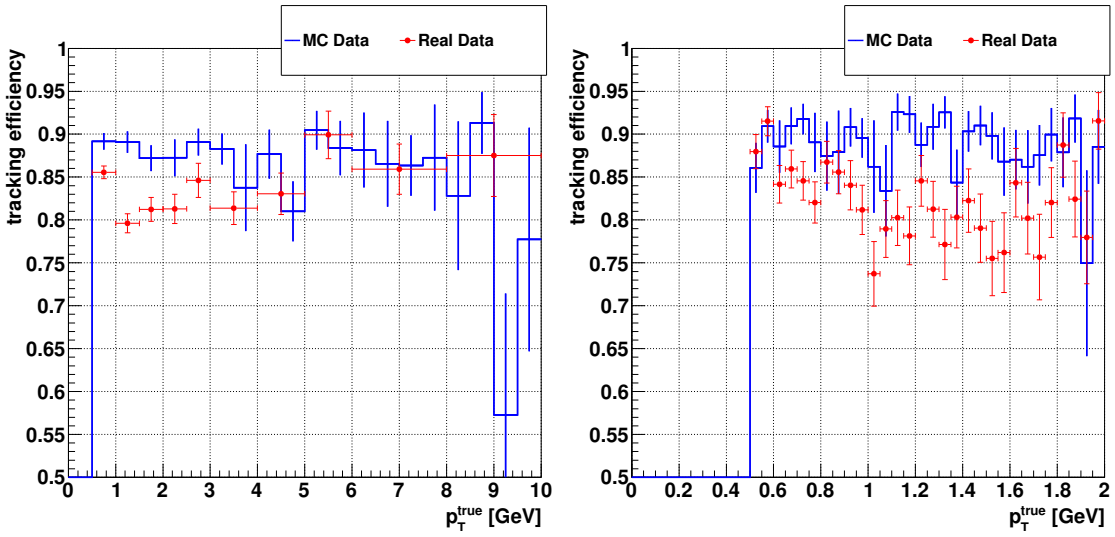


Fig. 20. Tracking efficiencies vs. p_T of embedded tracks, with tighter p_T and η cuts.

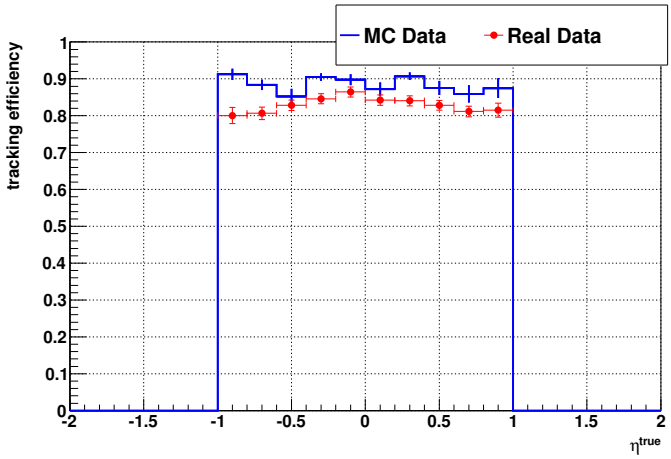


Fig. 21. Tracking efficiencies vs. η of embedded tracks, with tighter p_T and η cuts.

The figures indicate that at the p_T and η regions where the TPC has better tracking performance, the tracking efficiencies increase accordingly in both real data

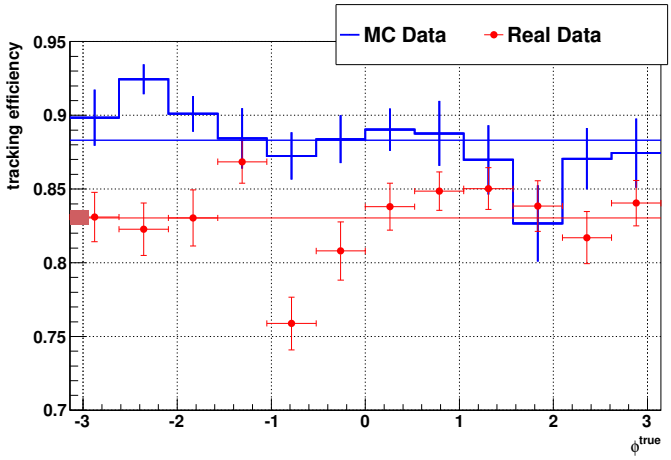


Fig. 22. Tracking efficiencies vs. ϕ of embedded tracks, with tighter p_T and η cuts.

embedding and MC embedding. However, the difference between the two embeddings are similar to the previous plots. Thus the calculation uncertainty is not significantly reduced when only considering conditions in which TPC has better performance. The average efficiences with the tighter cuts are summarized in Table V.

	Efficiency	Difference
Real Data Embedding	$83.0 \pm 0.5\%$	
MC Embedding	$88.3 \pm 0.6\%$	
		$5.3 \pm 0.8\%$

Table V. Comparison of tracking efficiencies with tighter p_T and η cut.

B. Tracking efficiency out of jet cones

Tracking efficiency differences between real data embedding and MC embedding for tracks lying out of jet cones were also analyzed, in order to check whether the dis-

crepancies come from the procedure of embedding a track within a specified jet-cone. Embedded tracks follow uniform distribution on 0.5 GeV $< p_T <$ 5 GeV, $-1 < \eta < 1$ and $0 < \phi < 2\pi$, ignoring the direction of the jet thrust axis. Since the area of jet cones in $\eta - \phi$ ($\pi \times 0.7^2 \simeq \pi/2$) is much smaller than the area where the embedded tracks were thrown (4π), most of the embedded tracks could be considered as lying out of jet cones. The results are shown in Figures 23, 24, 25, and Table VI. The discrepancy in the out-of-jet-cone tracking efficiency calculations, $6.1\% \pm 0.6\%$, is comparable to the figures for in-jet-cone tracking efficiencies.

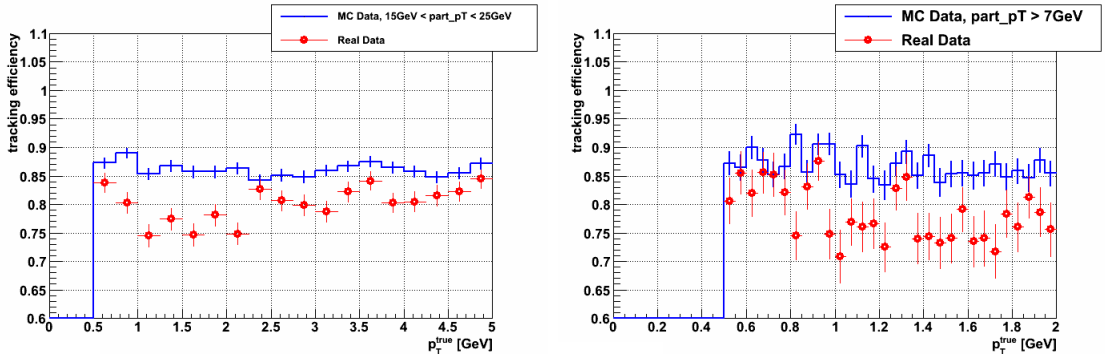


Fig. 23. Tracking efficiencies vs. p_T of embedded tracks, out of jet cones.

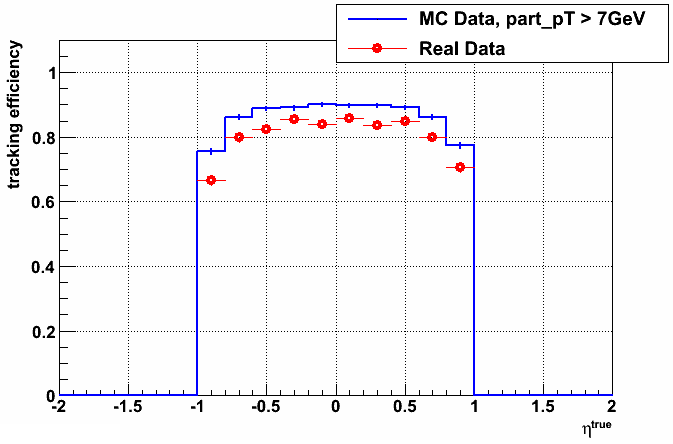


Fig. 24. Tracking efficiencies vs. η of embedded tracks, out of jet cones.

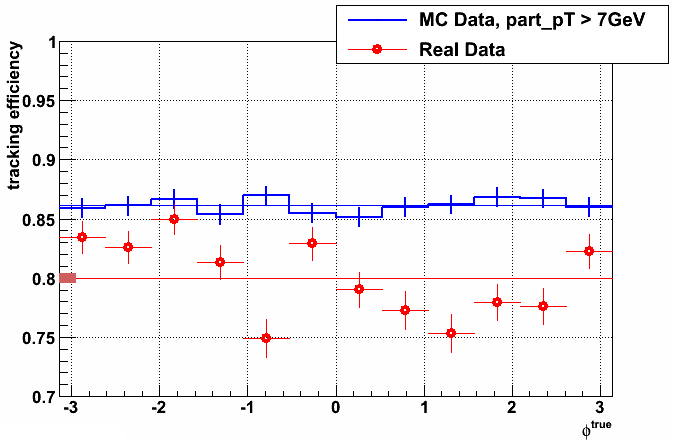


Fig. 25. Tracking efficiencies vs. ϕ of embedded tracks, out of jet cones.

	Efficiency	Difference
Real Data Embedding	$80.0 \pm 0.4\%$	
MC Embedding	$86.1 \pm 0.5\%$	
		$6.1\% \pm 0.6\%$

Table VI. Comparison of tracking efficiencies out of jet cones.

C. TPC gas gain fluctuation and effects for the 6% change of fudge factor

As discussed in Ch. III, Sec. B.2, the effect of the TPC gas gain fluctuation on tracking efficiency could be estimated by varying the "fudge factor" 6% in the real data embedding. Figures 26, 27 and 28 demonstrate how the tracking efficiencies will change when the fudge factor is varied by plus and minus 6% from the value of 1.26. The changes on the average tracking efficiency listed in Table VII are approximately 0.2%, which sit within statistical uncertainties and are well below the other efficiency differences found in this analysis.

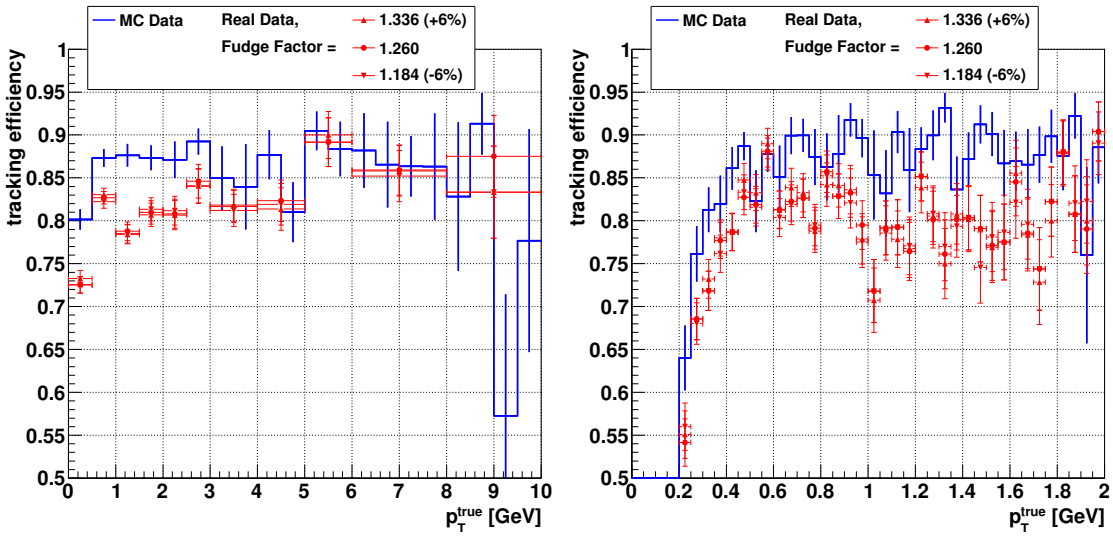


Fig. 26. Tracking efficiencies vs. transverse p_T of embedded tracks, with various fudge factors.

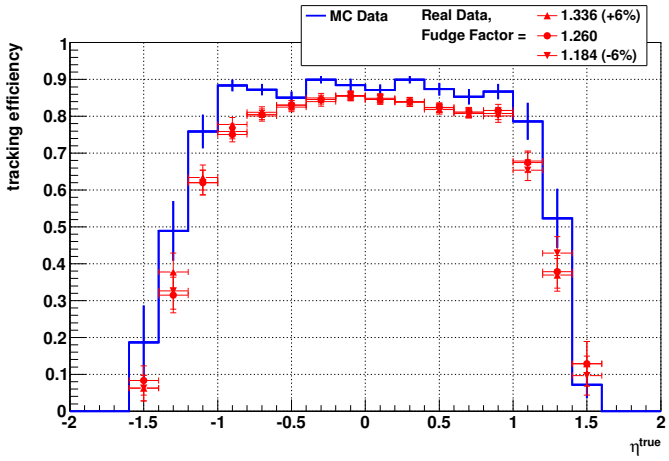


Fig. 27. Tracking efficiencies vs. transverse η of embedded tracks, with various fudge factors.

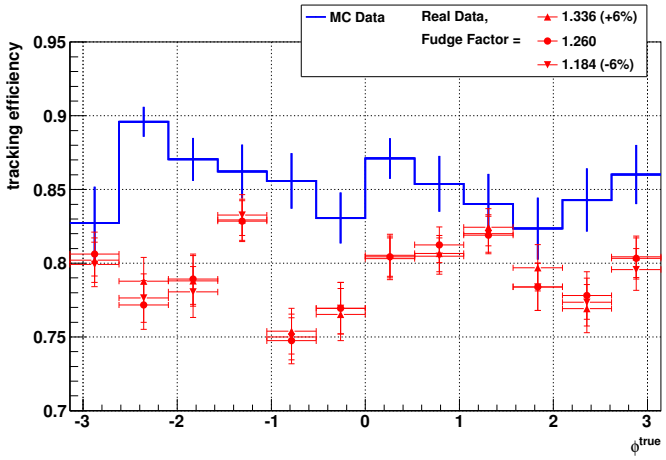


Fig. 28. Tracking efficiencies vs. transverse ϕ of embedded tracks, with various fudge factors.

	Fudge Factor	Efficiency	Difference*
Real Data Embedding	1.260	$79.5 \pm 0.4\%$	—
	1.336 (+6%)	$79.6 \pm 0.4\%$	$0.1 \pm 0.6\%$
	1.184 (−6%)	$79.3 \pm 0.4\%$	$-0.2 \pm 0.6\%$
MC Embedding		$85.4 \pm 0.5\%$	$5.9 \pm 0.6\%$

* Comparing to Real Data Embedding with Fudge Factor = 1.260.

Table VII. Comparison of tracking efficiencies with fudge factor variations.

D. Conclusion

In the presented analysis, TPC tracking efficiencies within reconstructed jet cones are calculated in both real data embedding and MC data embedding. Various factors that could affect the estimation of the tracking efficiency, such as the "fudge factor" in Monte-Carlo simulation, TPC gas gain fluctuations, and beam luminosity dependence, have been explored in this analysis. Post production cuts that are commonly used in jet analysis in proton-proton collisions were also taken into consideration.

In conclusion, the estimated tracking efficiency from embedding analysis of the STAR Run 6 longitudinal data is 79.5%. This figure represents the overall TPC tracking efficiency of tracks that are within reconstructed jet cones and fall into the kinematic region of $p_T > 0.2$ GeV and $-1.6 < \eta < 1.6$. The estimated systematic difference between the overall tracking efficiency obtained in real data embedding and MC data embedding is $5.9 \pm 0.6\%$(statistical). The presented analysis indicates that the difference is relatively independent on track kinematics, and is large enough to make the uncertainty introduced by the variation of TPC gas gain negligible. The 5.9% difference is consistent with the previously quoted $\sim 5\%$ systematic uncertainty used in standard STAR analysis [12]; however, it has been found in this analysis that this uncertainty could not be significantly reduced in the specific senario of jet analysis in polarized proton-proton collisions based on the existing Monte-Carlo event samples.

Though the origin of the discrepancy between real data embedding and MC data embedding could not be completely explained, the presented analysis suggests that a large portion of the difference is related to the high luminosity environment and the pile-up tracks. In MC data embedding, because the pile-up tracks were not taken into consideration and the resulted deterioration in track reconstruction was not reflected,

reconstructed embedded tracks have much better quality than those in real data embedding. Consequently, reconstructed embedded tracks in MC data embedding have a larger possibility to pass the cuts, which were commonly implemented in STAR jet analysis to ensure track reconstruction quality and eliminate pile-up tracks, than in real data embedding, as demonstrated in Figures 15 and 16. Table III indicates that these cuts introduced $\sim 2.7\%$ diference on tracking efficiency, which contributes $\sim 1/2$ to the calculated total systematic difference.

This portion of the difference could be addressed by introducing pile-up tracks to the MC embedding. Future analyses on simulated jets embedded into real zero-bias data should show little difference on the behaviors of the cuts in Table III between real data embedding and MC embedding combined with real zero-bias data. The remaining unexplained difference, by adding up the 3.2% difference[1] and 0.6% statistical uncertainty in quadrature, is up to 3.3%. Since the cause of this difference could not be identified, we adopt $\pm 3.3\%$ as the uncertainty on the tracking efficiency when embedding into real data under Run 6 conditions.

To further improve the accuracy on the estimation of the TPC tracking efficiency, more studies on the simulation tools used in this analysis are required in order to get a better understanding on how the high multiplicity environment from real data affect the embedded simulation track differently from the environment of pure simulation data, or from the environment of simulation data embedded into real zero-bias events. Also, much higher event statistics than were available from Run 6 will be required to permit a detailed investigation of the dpendence of the data–Monte-Carlo difference on track kinematics.

[1]Which comes from the total difference of 5.9%, minus the 2.7% from the cuts in Table III.

REFERENCES

[1] J. Blumlein and H. Bottcher. QCD analysis of polarized deep inelastic scattering data. *Nucl. Phys. B*, B841:205–230, 2010.

[2] Elliot Leader, Aleksander V. Sidorov, and Dimiter B. Stamenov. Determination of polarized PDFs from a QCD analysis of inclusive and semi-inclusive deep inelastic scattering data. *Phys. Rev. D*, 82:114018, 2010.

[3] Daniel de Florian, Rodolfo Sassot, Marco Stratmann, and Werner Vogelsang. Global analysis of helicity parton densities and their uncertainties. *Phys. Rev. Lett.*, 101:072001, 2008.

[4] Daniel de Florian, Rodolfo Sassot, Marco Stratmann, and Werner Vogelsang. Extraction of spin-dependent parton densities and their uncertainties. *Phys. Rev. D*, 80:034030, 2009.

[5] Gerry Bunce, Naohito Saito, Jacques Soffer, and Werner Vogelsang. Prospects for spin physics at RHIC. *Ann.Rev.Nucl.Part.Sci.*, 50:525–575, 2000.

[6] B. Jäger, M. Stratmann, and W. Vogelsang. Single-inclusive jet production in polarized pp collisions at $\mathcal{O}(\alpha_s^3)$. *Phys. Rev. D*, 70:034010, 2004.

[7] K. Ackermann et al. STAR detector overview. *Nucl. Instr. and Meth. A*, 499:624 – 632, 2003.

[8] M. Bai et al. Polarized proton collisions at 205 GeV at RHIC. *Phys. Rev. Lett.*, 96:174801, 2006.

[9] M. Harrison, T. Ludlam, and S. Ozaki. RHIC project overview. *Nucl. Instr. and Meth. A*, 499:235 – 244, 2003.

[10] M. Bai et al. RHIC beam instrumentation. *Nucl. Instr. and Meth. A*, 499:372 – 387, 2003.

[11] I. Alekseev et al. Polarized proton collider at RHIC. *Nucl. Instr. and Meth. A*, 499:392 – 414, 2003.

[12] M. Anderson et al. The STAR Time Projection Chamber: A unique tool for studying high multiplicity events at RHIC. *Nucl. Instr. and Meth. A*, 499:659, 2003.

[13] W.G. Gong. The STAR-TPC slow simulation. *STAR Note 197*, February 1995.

[14] M. Beddo et al. The STAR barrel electromagnetic calorimeter. *Nucl. Instr. and Meth. A*, 499:725–739, 2003.

[15] B.I. Abelev et al. Longitudinal double-spin asymmetry and cross section for inclusive jet production in polarized proton collisions at $\sqrt{s} = 200$ GeV. *Phys.Rev.Lett.*, 97:252001, 2006.

[16] B.I. Abelev et al. Longitudinal double-spin asymmetry for inclusive jet production in p+p collisions at $\sqrt{s} = 200$ GeV. *Phys.Rev.Lett.*, 100:232003, 2008.

[17] Gerald C. Blazey et al. Run II jet physics. *hep-ex/0005012*, 2000.

[18] Thomas Henry. *Reconstruction and attributes of jets observed in* $\sqrt{s} = 200$ *GeV proton-proton and deuteron-gold collisions at STAR*. PhD thesis, Texas A&M University, 2006.

[19] Monte carlo production datasets. `http://drupal.star.bnl.gov/STAR/comp/prod/monte-carlo-production-datasets`.

[20] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[21] P. Jacobs and D. Irmscher. GSTAR: A GEANT-based detector simulation chain for STAR. *STAR Note 235*, 1996.

[22] R. Brun, F. Bruyant, M. Maire, A.C. McPherson, and P. Zanarini. GEANT3. 1987. Revised version.

[23] Jeffery T. Michell and Iwona M. Sakrejda. Tracking for the STAR TPC: documentation and user's guide. *STAR Note 190*, 1994.

[24] `http://cyclotron.tamu.edu/star/2006Jets/sep9_2007/`.

[25] `http://www4.rcf.bnl.gov/~fisyak/star/PiD/RunVI/TpcGainVariationRunVI.png`.

[26] M. Sarsour. Constraints on Delta G through Longitudinal Double Spin Asymmetry Measurements of Inclusive Jet Production in Polarized p+p Collisions at 200-GeV. *AIP Conf.Proc.*, 1149:389–392, 2009.

[27] Benno List. Statistical error of efficiency determination from weighted events. `www.desy.de/~blist/notes/effic.ps.gz`.