Measurement Of Unpolarized Di-pion ($\pi^+\pi^-$) Cross-Section Using pp Collision Data At $\sqrt{s} = 200$ GeV At STAR

Babu Pokhrel (babu.pokhrel@temple.edu)
Bernd Surrow (surrow@temple.edu)
for the STAR collaboration

May 18, 2023

Contents

List of Figures

List of Tables

1 Introduction

A proton is a complex composite structure comprising three quarks: two up (u) and one down (d). Most of its spin structure comes from sources other than its valence quarks. After the compelling discovery of the EMC collaboration [1], understanding the proton structure has become a challenge for the particle physicist community. While the proton constituents are well understood, their contribution to the proton's spin is still puzzling. The valence quarks contribute about 30%, gluons can contribute up to 50%, and the rest should be coming from the orbital angular momentum of quarks and gluons, commonly called partons, in the total proton's spin.

At leading twist, a proton structure can be described by three parton distribution functions (PDFs): the unpolarized or helicity-averaged ($f_1(x)$) PDF, the helicity PDF ($g_1(x)$), and the transversity PDF ($h_1(x)$), where x is the fractional momentum of the proton carried by the parton. The first two PDFs are well understood from the semi-inclusive deep inelastic scattering (SIDIS) experiment; however, the latter has very limited knowledge to date. Recently, $h_1(x)$ has gained attention due to its correlation with the nucleon tensor charge (δq), an important nucleon property. However, measuring $h_1(x)$ is not simple, as it is related to the helicity flip mechanism, which is suppressed in the DIS process. $h_1(x)$ appears only in processes where a hadron is involved, at least one, directly in the initial or final state. Moreover, being chiral odd, $h_1(x)$ must be coupled to another chiral-odd quantity to be observed. A proton-proton (pp) collision would be the ideal condition for the extraction of $h_1(x)$, where detecting hadrons within jets or different hadronic final states allows $h_1(x)$ to be coupled with chiral-odd fragmentation functions (FFs). For example, $h_1(x)$ can be coupled with the Collins FF, $H_1^{\perp}(z, j_t)$, if we detect hadrons within jets, or it can be coupled with the di-hadron interference FF, $H_1^{\triangleleft}(z, M)$, if we detect di-hadron pairs in the final state.

An observable, dihadron correlation asymmetry (A_{UT}) for pions, has been measured by the STAR collaboration [2, 3, 4], which is sensitive to the product of $h_1(x)$ and $H_1^{\triangleleft}(z, M)$. $h_1(x)$ can be singled out from the product by measuring the interference fragmentation function (IFF) independently in the e^+e^- process, thanks to the factorization and universality of the physics mechanism, at least in the collinear framework. However, one crucial observable that needs to be experimentally measured is the unpolarized dihadron cross-section ($d\sigma_{UU}$), which is necessary for the extraction of the unpolarized dihadron FF (D_1). The experimental results on the $d\sigma_{UU}$ are very limited, and the parameterization of D_1 relies on the simulation resulting in a large uncertainty.

This article presents the initial measurement of the unpolarized $\pi^+\pi^-$ cross-section in pp collisions at the center-of-mass energy ($\sqrt{s}$) of 200 GeV using data from the STAR Run 2012. The dataset is dominated by the jet patch triggered with a lower energy threshold than the run 2015 dataset, making it a suitable choice for measuring the cross-section as it enhances the sensitivity to gluons. **********

2 Analysis

2.1 Dataset

The STAR run 12 pp collision data at $\sqrt{s} = 200$ GeV is used in this analysis. This particular dataset is chosen as it comprises triggered events of lower energy threshold, which provides higher gluon sensitivity.

2.2 PYTHIA Simulation And Embedding

To produce simulated particles for the proton-proton collisions at the center-of-mass energy $\sqrt{s} = 200$ GeV, known as "particle level", we use the PYTHIA 6 Monte Carlo event generator [5]. For this purpose, we rely on a Perugia 12 [6] with the modified tune (PARP(90)=0.213) [7], which utilizes the CTEQ6 PDF sets [8].

Next, we simulate the response of the STAR detector using GEANT3 [9]. Before the actual reconstruction is performed, we mix the raw simulated detector responses event-by-event with the actual detector responses from the Zerobias trigger sample, which is called the embedding process. For the embedding, we only use events sampled by the Zerobias trigger since it represents random detector states.

More information regarding the production of the embedding sample can be found in the STAR thesis [10].

2.3 Event Selection

The event and track selection criteria employed by the Spin group for this analysis are standard. The primary vertex with the highest rank ($V_{z,rank} > 10e6$) must be within 60 cm from the center of the Time Projection Chamber (TPC) along the beam direction. In addition, only events satisfying the JP0, JP1, or JP2 triggering criteria are used, as these are the dominant triggers in the data sample, which are selected using the offline trigger IDs 370601, 370611, and 370621 for the JP0, JP1, and JP2, respectively. Primary tracks from the selected primary vertices that pass track level quality cuts are used to identify pions, which is based on the ionization energy loss as the charge particles traverse through TPC gas volume (dE/dx). A cut of $-1 < n\sigma_\pi < 2$, where $n\sigma_\pi$ is the quantity calculated using dE/dx, and the theoretical model, is used as the default pion selection cut throughout the analysis.

Pairs of oppositely charged pion tracks within a collision event are used to construct all-inclusive $\pi^+\pi^-$ events. $\pi^+\pi^-$ events with opening angles in $\eta - \phi$ space greater than 0.7 or smaller than 0.02 are rejected. The upper cone cut is intended to ensure that the $\pi^+\pi^-$ is within the jet; note that the jet reconstruction is not required in this channel. The minimum cone cut is determined from a study of the cone resolution using embedding, which excludes $\pi^+\pi^-$ events formed from tracks that are very close that the detector cannot identify as separate. The cone resolution, a difference between the $\pi^+\pi^-$ invariant mass at the particle and detector level, for different upper cone cuts, is shown in figure 1. The cone resolution is considered as the Gaussian width of the mass difference at the particle and detector level, which has no upper cone cut dependence. The cone resolution is found to be 0.018 GeV/c^2, and the minimum cone cut greater than 0.02 GeV/c^2 is applied throughout this analysis. Additionally, hard cuts are applied to the pair p_T ($1 < p_T^{\pi^+\pi^-} < 15$ Gev/c) and pair mass ($0.27 < M_{inv}^{\pi^+\pi^-} < 4(GeV/c^2)$) to further refine the selection of $\pi^+\pi^-$ events.

All the selection cuts applied in this analysis are listed in the table 1. In the table, "Detector" refers to the reconstructed quantities, "particle" refers to the corresponding Monte-Carlo generated PYTHIA quantities associated with the detector quantities, and "PYTHIA" refers to the pure PYTHIA particles.
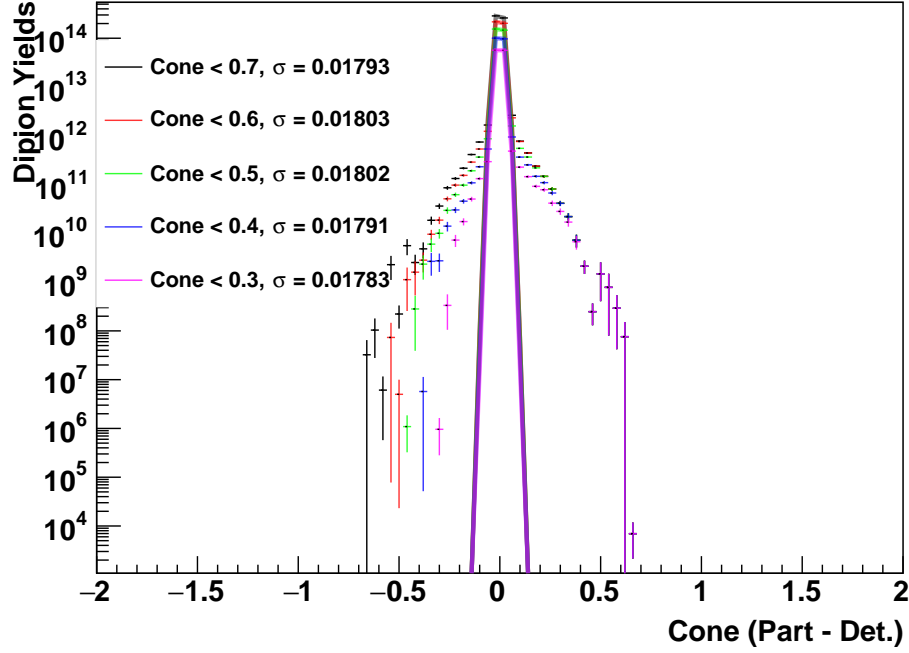
Figure 1: **Cone resolution** - Difference between generated and reconstructed $\pi^+\pi^-$ invariant mass for different upper cone cut.

Quantities	Selection Cuts		
	Data	Detector	Particle (PYTHIA)
Z-Vertex(V_z, cm)	< 60	< 60	
$V_{z,rank}$	$> 1e6$	$> 1e6$	
Trigger	JP0, JP1, JP2	JP0, JP1, JP2	
$p_{T,track}(GeV/c)$	$0.5 < p_T < 15$	$0.5 < p_T < 15$	$0.5 < p_T < 15$
$|\eta_{track}|$	< 1	< 1	< 1
$dca(cm)$	p_T dependence	p_T dependence	
nHitsFit	> 15	> 15	
nHits Ratio	> 0.51	> 0.51	
Pion Selection	$-1 < n\sigma_\pi < 2$	$-1 < n\sigma_\pi < 2$	GePID:8,9 (pdg:$\pm$211)
Cone	$0.02 < Cone < 0.7$	$0.02 < Cone < 0.7$	$0.02 < Cone < 0.7$
$\eta^{\pi^+\pi^-}$	$|\eta^{\pi^+\pi^-}| < 1$	$|\eta^{\pi^+\pi^-}| < 1$	$|\eta^{\pi^+\pi^-}| < 1$
$p_T^{\pi^+\pi^-}(GeV/c)$	$1 < p_T^{\pi^+\pi^-} < 15.0$	$1 < p_T^{\pi^+\pi^-} < 15.0$	$1 < p_T^{\pi^+\pi^-} < 15.0$
$M_{inv}^{\pi^+\pi^-}(GeV/c^2)$	$0.27 < M_{inv}^{\pi^+\pi^-} < 4.0$	$0.27 < M_{inv}^{\pi^+\pi^-} < 4.0$	$0.27 < M_{inv}^{\pi^+\pi^-} < 4.0$
$cone = \sqrt{(\eta^{\pi^+} - \eta^{\pi^-})^2 + (\phi^{\pi^+} - \phi^{\pi^-})^2}$			

Table 1: Event and track selection cuts

The dca has the following p_T dependence:

$$\text{dca} = \begin{cases} < 2 \text{ cm}, & \text{if } p_T < 0.5 \text{ GeV} \\ < 2.5 - \ p_T \ \cdot (1 \ \text{cm/GeV}), & \text{if } 0.5 \leq p_T < 1.5 \text{ GeV} \\ < 1 \text{ cm}, & \text{if } p_T \geq 1.5 \end{cases}$$

To measure the unpolarized $\pi^+\pi^-$ cross-section, it is necessary to bin the data by invariant mass (and in the p_T would be more impactful, though this analysis only considers mass bins). The width of the bins must be greater than the invariant mass resolution, which is derived from the individual momenta of the pion tracks and their energies in each pair. The energy of the identified pions is calculated using known pion mass ($M^{\pm\pi} = 0.1396$ GeV/c^2) and the pion's momentum from the TPC. To study the resolution, an embedding sample is used that matches the data well. The difference between the $\pi^+\pi^-$ mass at the detector level and at the particle level is considered as the mass resolution, which is studied in 12 mass bins with approximately equal widths of 0.3 GeV/c^2. Figure 2 shows the differences in the detector and particle masses for each bin, and the Gaussian width of each distribution is considered the mass resolution. The bin resolution is plotted as vertical error bars as a function of the average mass in Figure 3. The resolution is better in the lower mass region (< 0.03) than in the higher mass region. This mass-resolution dependence is consistent with the (transverse) momentum resolution of the particle tracks as the invariant mass has a linear relation to the transverse momentum of the $\pi^+\pi^-$.

2.4   Cross Section Binning
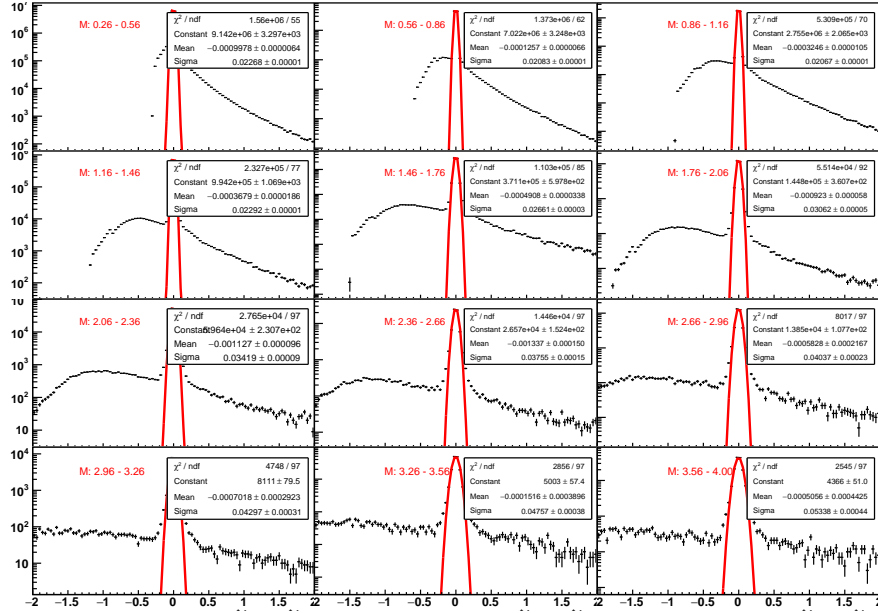
Figure 2: $M_{gen}^{\pi^+\pi^-} - M_{rec}^{\pi^+\pi^-}$ distributions in 12 mass bins each of 0.30 GeV/c^2 fitted with the *gaussian* function.

8
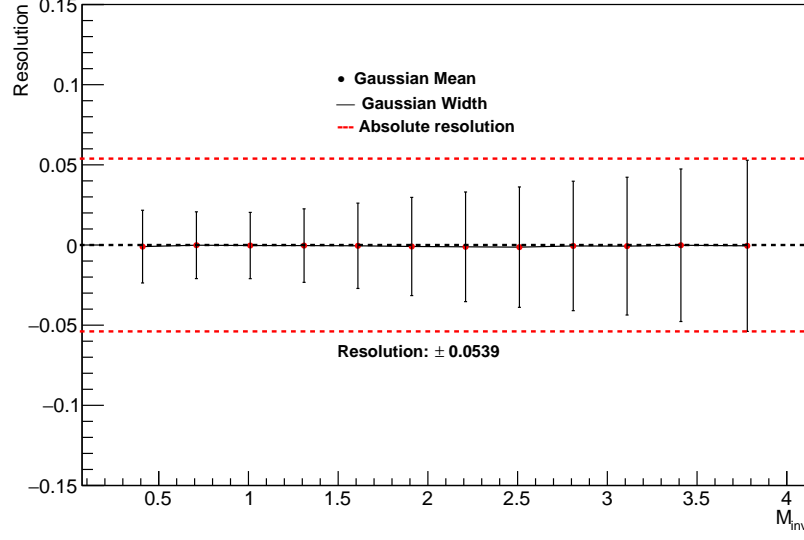
Figure 3: Bin resolution (gaussian width from figure 2) as a function of dipion mass. The envelope of the vertical error bars gives the absolute resolution (bin resolution when mass integrated.)
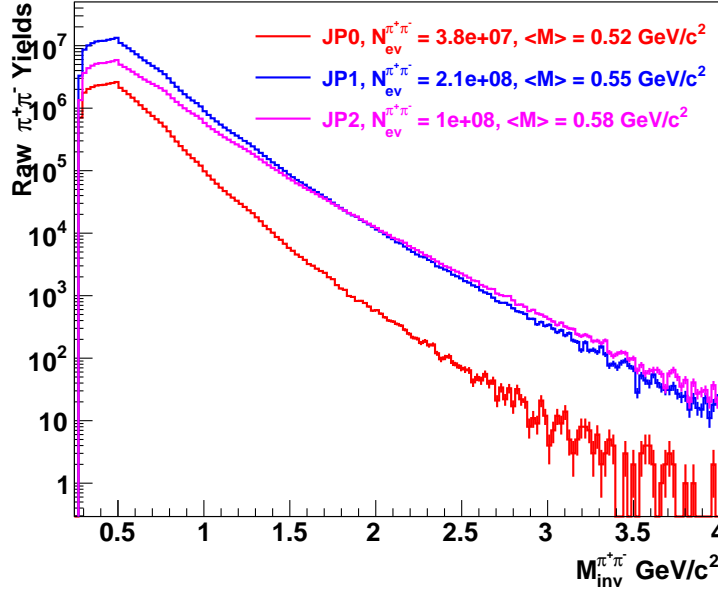
Figure 4: **Raw $\pi^+\pi^-$ yields** - Raw yields for JP0, JP1, and JP2 triggers.

The bin width for the cross-section measurement of the $\pi^+\pi^-$ invariant mass should not be smaller than the bin resolution. While the bin resolution varies with the mass, a minimum bin width is chosen to be greater than the absolute resolution, which is $0.054\ GeV/c^2$. This ensures that the bin width is large enough to avoid any smearing effects on the measured cross-section.

The $\pi^+\pi^-$ yields for each trigger are shown in Figure 4. The dominant statistics come from the JP1-triggered events, followed by JP2 and JP0, respectively. The yields are lower at higher masses for JP0 than for JP1 and JP2 due to the trigger threshold energy. For all triggers, the same mass

9

binning is used, with the width varied to account for the decreased yields at higher masses. The selected mass bins for the cross-section and other relevant kinematics are shown in Table 3.

2.5 Luminosity

In the STAR experiment, events are selected based on specific triggering conditions to meet particular physics goals. This analysis uses three jet-patch triggers: JP0 (ID = 370601), JP1 (ID = 370611), and JP2 (ID = 370621), each with a different trigger threshold, resulting in different luminosity per trigger. The luminosities per trigger are recorded during the data-taking process and are available online (JP0, JP1, JP2). Luminosity data per run for JP0, JP1, and JP2 are read and added from these files to obtain the total luminosity, which is calculated for 601 good physics runs that are used in this analysis (run list in the appendix C). Also, the average prescale value for each trigger is calculated from the total number of runs used by reading values per run from the same files.

Triggers	JP0	JP1	JP2	
Trig. ID	370601	370611	371621	$L_{JP} = 26.64 \ pb^{-1}$
$L(pb^{-1})$	0.16	7.68	18.80	
Prescale	118.335	2.514	1	

Table 2: Total luminosity and prescale value per trigger.
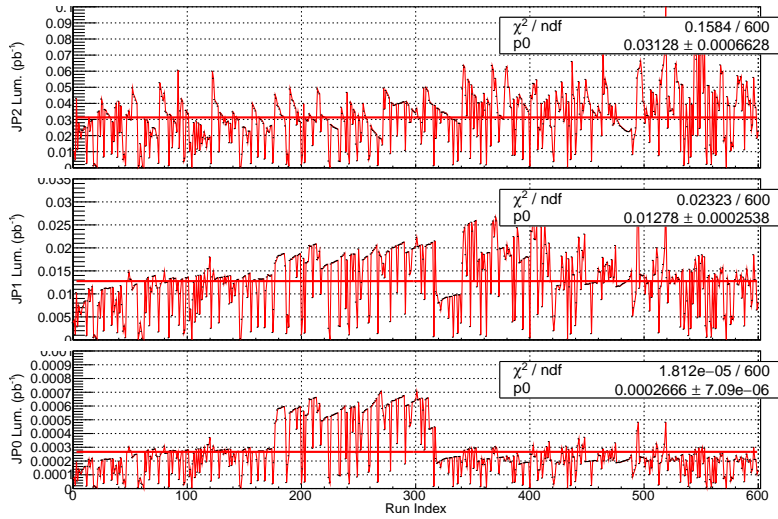
Figure 5: **Luminosity vs Run index** - Luminosity as a function of the run index, where the run index is assigned by sorting run numbers in increasing order, for JP0 (*bottom*), JP1 (*middle*), and JP2 (*top*). Luminosities per trigger per run for this dataset are obtained available online (JP0, JP1, JP2).

Table 2 shows the total luminosity and average prescale values for each trigger for the whole data sample used in this analysis. Figure 5 displays the recorded luminosity as a function of the run index for JP0 (bottom), JP1(middle), and JP2 (top) triggers. The saw-tooth pattern in the plot depicts the fill dependence, where the recorded luminosity falls off with time within the fill.

2.6 Data Unfolding

The unfolding analysis is performed using the TUnfoldDensity algorithm provided in the ROOT package. The unfolding procedure unfolds the measured data spectrum to the true level, which can be achieved using a Monte-Carlo simulated embedding sample, which should have a very good description of the measured spectrum. An officially produced embedding sample for this dataset (embedding request page), which has a very good description of the data, is used in this analysis. A detailed comparison between the data and the embedding sample can be found in the appendix (see B).

2.6.1 $\pi^+\pi^-$ background definition and normalization scale

The embedding saves information about the generated ("generated" and "true" refers to the same PYTHIA generated quantities recorded by GEANT, which are used interchangeably in this document), and the reconstructed events and tracks quantities. There may have been some particle tracks reconstructed and stored which are not associated with any of the generated tracks. Those are the background tracks, which can be identified via the idTruth of a reconstructed track, which gives information about the association between the generated and reconstructed tracks. Any of the reconstructed tracks with the zero idTruth value are not associated with the generated tracks, and they should be treated as a background. In addition, the matching condition is imposed between the idTruth of the reconstructed and the track ID of the generated for the certainty of the association. Thus, to consider a signal track, it has to match both conditions, which is otherwise considered a background track. The definition of a background at the $\pi^+\pi^-$ level is based on whether a pair consists of the background track or not. In a $\pi^+\pi^-$ pair, if any or both of the tracks are background, those pairs are considered background pairs. Figure 6 shows the raw $\pi^+\pi^-$ yields from the data (shown in red), backgrounds from the embedding (shown in blue), and background-subtracted data (shown in green) for JP0, JP1, and JP2 triggers from left to right, respectively. The $\pi^+\pi^-$ background yield from the embedding, which contributes only in the lower mass region, is scaled so that it can be directly compared to the data. The scale factor, S_{bkg}, is determined from the total $\pi^+\pi^-$ yields in the data (I_{dat}), and the total $\pi^+\pi^-$ yields in the embedding, I_{emb}, as,

$$S_{bkg} = \frac{I_{dat}}{I_{emb}} \tag{1}$$

2.6.2 Binning for the unfolding

Different unfolding algorithms may have different binning requirements; otherwise, the unfolding results wouldn't be trustworthy. For most unfolding algorithms, there are at least two inputs: the input data spectrum to be unfolded and the response matrix, which is a two-dimensional matrix with the "true" values on one axis and corresponding reconstructed values on the other. Whatever the axis choice, unfolding should always be performed on the "true" axis. For proper unfolding using the TUnfoldDensity algorithm, which is used in this analysis, the response matrix must have the number of "reconstructed" bins (n) greater than or equal to the number of "true" bins (m): $n \geq m$. The data input spectrum must match the number of reconstructed bins in the response matrix. This bin number criterion is crucial; however, the choice of bin width does not affect the reliability of the unfolding output.
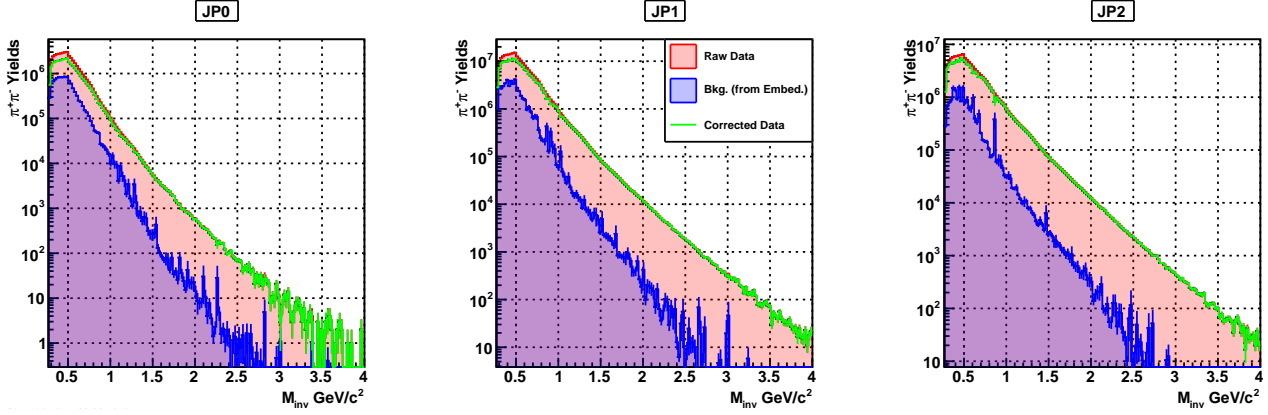
Figure 6: $\pi^+\pi^-$ **signal and background yields** - Total $\pi^+\pi^-$ yields from the data ("Raw Data" in red), $\pi^+\pi^-$ backgrounds from embedding ("Bkg." in blue), and background subtracted data ("Corrected Data" in green) as a function of $\pi^+\pi^-$ invariant mass, $M_{inv}^{\pi^+\pi^-}$, for the JP0, JP1, and JP2 triggers from left to right, respectively. Backgrounds are multiplied by the scale factor S_{bkg}.

Bin Id	M-Range(GeV/c^2)	BW	$< M_{inv} > (GeV/c^2)$	$< p_T^{\pi^+\pi^-} > (GeV/c)$	$< \eta^{\pi^+\pi^-} >$
1	0.27 - 0.35	0.08	0.319	1.74	-0.0168
2	0.35 - 0.45	0.1	0.401	1.87	-0.0173
3	0.45 - 0.6	0.15	0.517	2.06	-0.0176
4	0.6 - 0.75	0.15	0.666	2.64	-0.0197
5	0.75 - 0.95	0.2	0.83	3.42	-0.0217
6	0.95 - 1.15	0.2	1.03	4.3	-0.0233
7	1.15 - 1.35	0.2	1.24	5.22	-0.0243
8	1.35 - 1.6	0.25	1.45	6.16	-0.0264
9	1.6 - 1.9	0.3	1.72	7.24	-0.0252
10	1.9 - 2.2	0.3	2.02	8.33	-0.0273
11	2.2 - 2.6	0.4	2.36	9.33	-0.0289
12	2.6 - 3.2	0.6	2.81	10.5	-0.0337
13	3.2 - 4	0.8	3.45	11.9	-0.0149

Table 3: Binning and kinematics for the $\pi^+\pi^-$ cross-section.

We constructed the response matrix of $n = 160$ uniform bins and $m = 13$ variable-width bins, which correspond to the cross-section bins (see details in section 2.4) onto which the data spectrum is unfolded. The cross-section binning and the relevant kinematic values in those bins are presented in the table 3. The optimum value of n is selected after several iterations for a reasonable unfolding output. The response matrices for JP0, JP1, and JP2 triggered events are shown in the top-left panel of figures 7-9, respectively, where the y-axis contains "true" values and the x-axis contains "reconstructed" values of $M_{inv}^{\pi^+\pi^-}$. The response matrix shows a good correlation between the true and reconstructed masses, as most of the entries fall along the diagonal.

2.6.3 Unfolding using TUnfoldDensity algorithm

The TUnfoldDensity algorithm [11] effectively handles background through the SubtractBackground() method, which takes the background histogram and normalization scale factor as input.

Unfolding is performed with the background subtraction for each of the triggers individually, which allows us to measure the cross-section for each trigger. The $\pi^+\pi^-$ invariant mass, which is to be unfolded, and background distributions are shown in figure 6.

Figures 7-9 show the unfolding analysis details for triggers JP0, JP1, and JP2, respectively. The top right panel displays the L-curve, which is a scanned region for the regularization parameter, τ, and the asterisk shows the minimum curvature of the scanned region, which is used to calculate the τ. For good unfolding results, the L-curve should have an approximate L-shape with the kink somewhere on the scanned region but not too close to the edges. The bottom left panel presents the bin correlation matrix between the generated (y-axis) and reconstructed (x-axis) bins. A good bin correlation matrix should be diagonal.
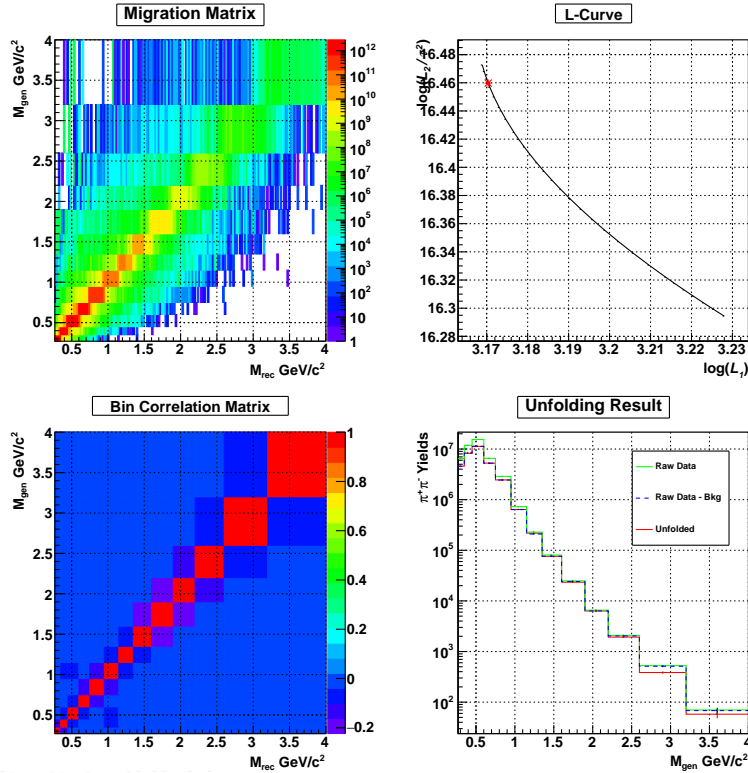
Figure 7: **JP0 Unfolding** - Unfolding output for the JP0 triggered events.

The bottom right panel depicts the raw data, the background-subtracted data, and the unfolded distribution into 13 cross-section bins. The input data histograms for the unfolding have a different number of bins than those shown here; specifically, the data input consists of 160 reconstructed bins of uniform width, while the output bins are 13 cross-section bins at the truth level. As expected, the unfolded output differs slightly from the input, with yields from the higher mass bins migrating to the lower mass region. This final unfolded distribution is devoid of the bin migration effect and background and is at the true level. However, the unfolded yields still require several corrections that are discussed in the following sections.

2.7 Efficiency Studies

Two types of efficiencies appear in this analysis; tracking and triggering efficiencies. In high-energy physics, these types of analyses are performed using the embedding sample as it has both generated
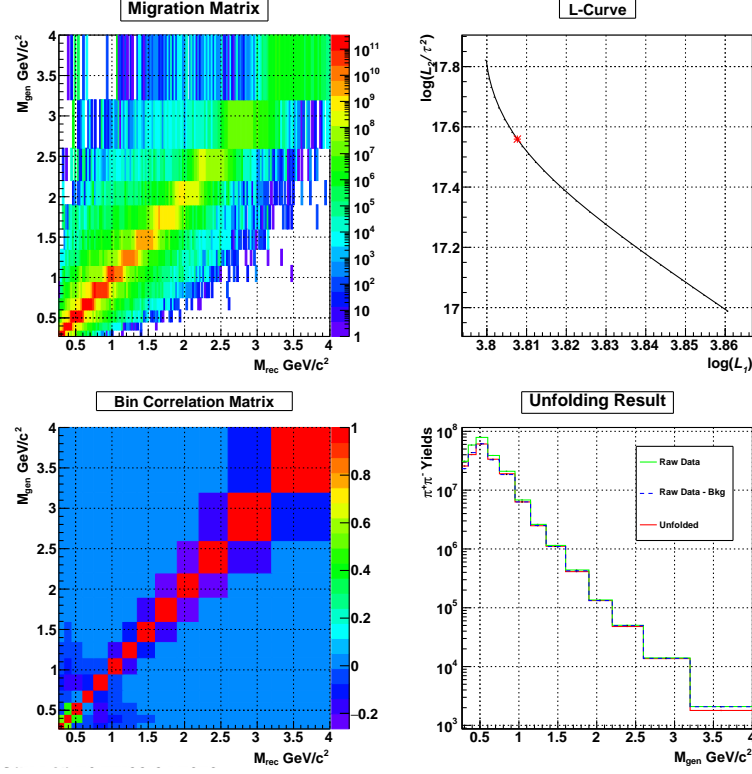
Figure 8: **JP1 Unfolding** - Unfolding output for the JP1 triggered events.
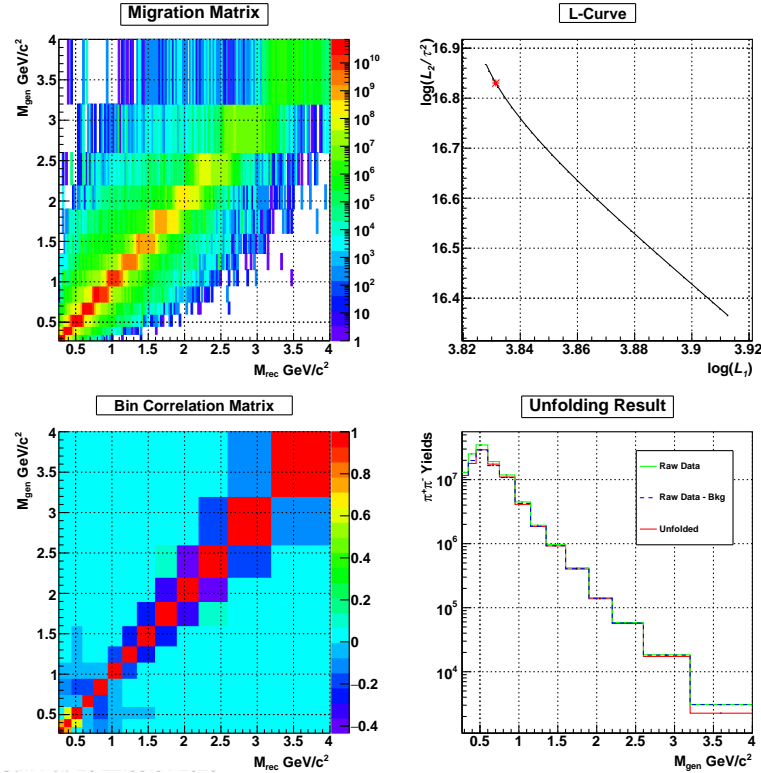
Figure 9: **JP2 Unfolding** - Unfolding output for the JP2 triggered events.

and reconstructed event information. Each of these analyses is done separately in the following sub-sections. For all the efficiency analysis, the detector level cuts as listed in the table 1 are applied. Only the p_T and η cuts are retained at the particle level for the consistency of the kinematic and acceptance effect in both cases.

2.7.1 Tracking Efficiency (ϵ_{trk})

Mathematically, the tracking efficiency (ϵ_{trk}) is defined as the ratio of the number of reconstructed tracks to the number of generated tracks;

$$\epsilon_{trk}^{\pi} = \frac{N_{gen,rec}^{\pi}}{N_{gen}^{\pi}} \tag{2}$$

Where N_{gen} is the number of generated tracks and $N_{gen,rec}$ is the number of generated tracks that passed all the detector level cuts. For the physics analysis, tracks must satisfy certain quality cuts. Both generated and reconstructed tracks are subjected to common event-level selection cuts of $|V_z| < 60$ cm, $V_{z,rank} > 1e6$ and additional cuts on track p_T and η (*common cuts*). In addition to the common cuts, the track is considered reconstructed if it passes all the detector cuts listed in the table 1. All the selection cuts are applied at the detector level. Only the fiducial p_T and η cuts are retained at the particle level tracks.

Figure 10 illustrates the tracking efficiency for π^+, π^-, their product, and the $\pi^+\pi^-$ efficiency, as a function of the $\pi^+\pi^-$ invariant mass, for each trigger used. The top panel displays the tracking efficiency for π^+ (left) and π^- (right), in cross-section $M_{inv}^{\pi^+\pi^-}$- bins. Pion tracking efficiency increases with mass in the lower mass region but tends to decrease in the higher mass regions. The bottom left panel highlights the mass dependence of the tracking efficiency for a $\pi^+\pi^-$, which is the product of pion track efficiencies in a pair. The bottom right panel shows the tracking efficiency at the $\pi^+\pi^-$ level, which is consistent with the product-level efficiency and provides a consistency check of the two approaches. The product-level efficiency is applied for the correction in each bin of the cross-section.

2.7.2 Triggering Efficiency (ϵ_{trg})

Trigger efficiency correction is necessary for cross-section measurement analysis in particle physics because the cross-section is a measure of the probability of a certain interaction occurring, and the trigger system plays a crucial role in selecting the events of interest. The trigger system is designed to select only a subset of events that are of interest for a particular analysis, given the high rate of events produced from the collisions. The trigger efficiency correction aims to account for these inefficiencies and to ensure that the observed number of events in the selected sample accurately reflects the number of events that would have been observed in the absence of the triggering condition.

The triggering efficiency can be estimated from the embedding sample, as embedding doesn't know about the event triggering unless we invoke the trigger simulator algorithm. Having both the event information, with and without the event triggering effects, we can calculate the triggering efficiency for a particular trigger of interest. The ϵ_{trg}, for a particular trigger, is defined as the ratio of the number of events triggered to the number of unbiased events. Mathematically,

$$\epsilon_{trg} = \frac{\sum N_{triggered}^{\pi^+\pi^-}}{\sum N_{unbiased}^{\pi^+\pi^-}} \tag{3}$$
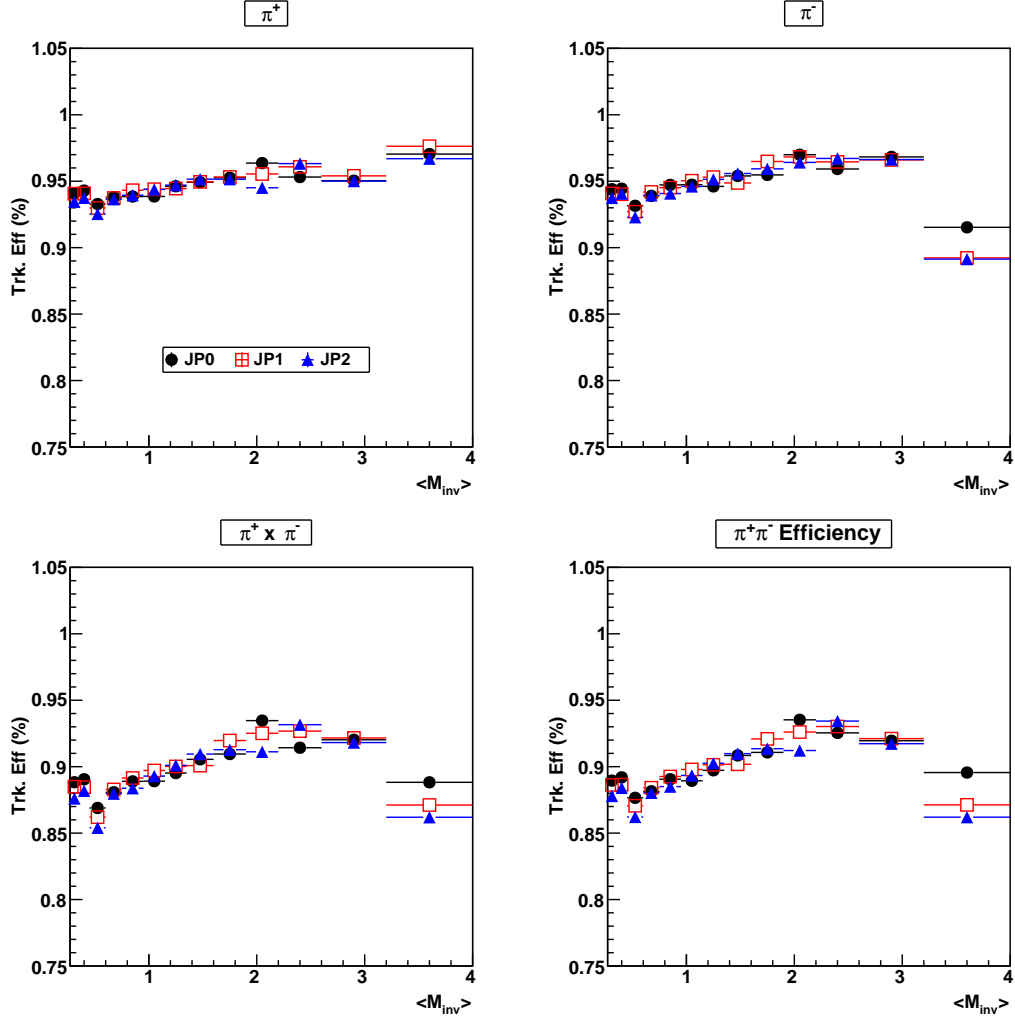
Figure 10: ϵ_{trk} **vs** M_{inv}. - Tracking efficiencies in the cross-section $M_{inv}^{\pi^+\pi^-}$- bins are shown for JP0, JP1, and JP2 triggers. The final $\pi^+\pi^-$ tracking efficiency is the product of the pion track efficiencies in a pair, shown in the bottom left panel. The bottom right panel is the $\pi^+\pi^-$ tracking efficiency, which is consistent with the product of the pions track efficiencies.

where, $N_{triggered}^{\pi^+\pi^-}$ represents the number of $\pi^+\pi^-$ events from a triggered event, while $N_{unbiased}^{\pi^+\pi^-}$ represents the number of $\pi^+\pi^-$ events without the application of a trigger algorithm, both at the truth level. The "truth level" event refers to the $\pi^+\pi^-$ event from the Monte-Carlo tracks associated with the reconstructed pion tracks, with all the selection cuts applied at the detector level. The trigger efficiency is calculated for each invariant mass bin for the JP0, JP1, and JP2 triggers using Equation 3, as shown in Figure 11. The triggering efficiency is lower in the lower mass region and higher in the higher mass region for all triggers. The JP0 trigger is activated more frequently, as it has the lowest energy threshold, while the JP1 and JP2 triggers are activated less often due to their higher energy threshold criteria. The trigger efficiency correction is applied for each of the cross-section bins.
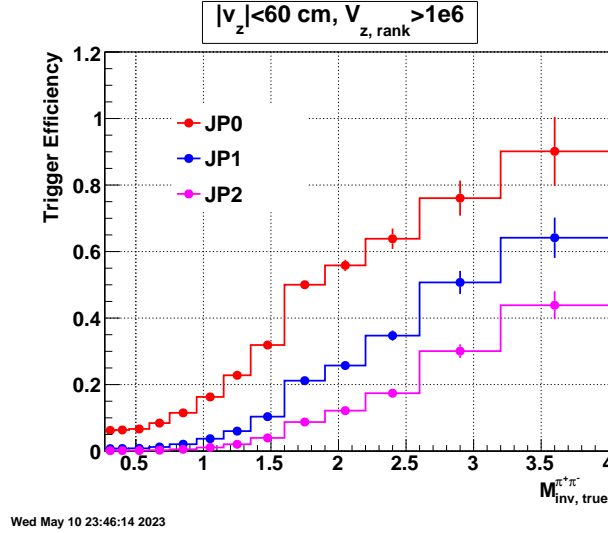
Figure 11: ϵ_{trg} **vs** $M_{inv}^{\pi^+\pi^-}$ - Trigger efficiency for JP0, JP1, and JP2 triggers as a function of $M_{inv}^{\pi^+\pi^-}$ in cross-section bins.

2.8 Corrections for misidentification and loss of pions

This type of correction is usually done with the data-driven approach; however, we have used an embedding-driven approach as there is a good match of $n\sigma_\pi$, which is used for the particle identification (PID), between the data and embedding (see section B). For the final publication, the consistency between the data-driven and this approach will be studied.

We have considered two corrections related to the PID: one is associated with the misidentification of protons and kaons as pions within the default $n\sigma_\pi$ cut. This tends to increase the $\pi^+\pi^-$ yields because of the fake pion contribution. The other correction is associated with the loss of true pions because of the default $n\sigma_\pi$ cut, which needs to be kept. This accounts for the inefficiency of the default pion selection adopted in this analysis. Each of these corrections is discussed in the following sections.

2.8.1 Correction for the fake pion contribution (f_{fake})

The f_{fake} is calculated as a fraction of true $\pi^+\pi^-$ pairs within the default $n\sigma_\pi$ cut, which accounts for the fake pions that enter into the signal region. Mathematically,

$$f_{fake} = \frac{N^{pair}_{n\sigma\pi\&\&true}}{N^{pair}_{n\sigma\pi}} \tag{4}$$

where the denominator is the $\pi^+\pi^-$ pair selected with the default $n\sigma_\pi$ cut at the detector level, and the numerator is the number of detected pairs that are true pairs (true pairs are identified by the GEANT particle ID: 8 for π^+ and 9 for π^-).
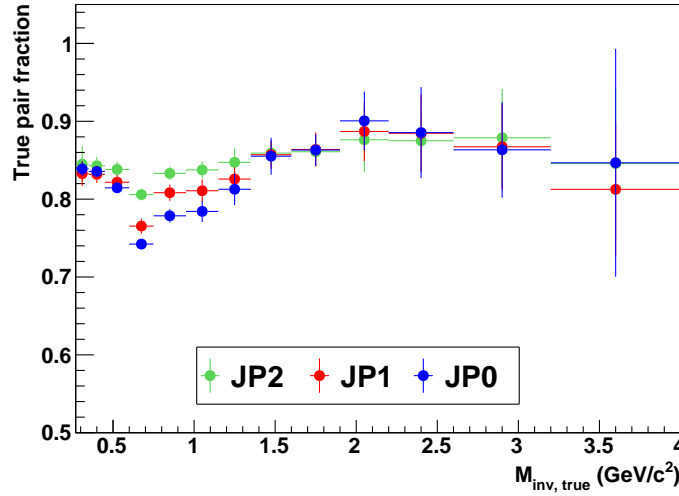
Figure 12: f_{fake} - Fraction of true pion pair. This fraction accounts for the fake pion contribution in the $\pi^+\pi^-$ yields; hence the "fake" subscript is introduced in the symbol.

Figure 12 shows the f_{fake} in $M^{\pi^+\pi^-}_{inv}$ cross-section bins for all three triggers, JP0, JP1, and JP2. f_{fake} has a mild mass dependence, with the increase in the fraction as the mass increases. Also, the fraction has a trigger dependence, specifically in the region $M^{\pi^+\pi^-}_{inv} < 1.5\ GeV/c^2$; the fraction is lower for the trigger of a lower trigger threshold. The correction for fake contributions is applied at the triggered cross-section level on a bin-by-bin basis.

2.8.2 Correction for the pion loss (f_{loss})

Some of the true $\pi^+\pi^-$ may be lost due to the restrictive PID cut. This loss of $\pi^+\pi^-$ events can be estimated by calculating the fraction;

$$f_{loss} = \frac{N_{true}}{N_{n\sigma\pi\&\&true}} \tag{5}$$

where N_{true} is the number of true $\pi^+\pi^-$ pairs and $N_{n\sigma\pi\&\&true}$ is the number of true $\pi^+\pi^-$ pairs which falls within the $-1 < n\sigma_\pi < 2$ range. This accounts for the inefficiency of the applied PID cut on identifying pion. This fraction is greater than 1 and increases the yields by this factor to account for the loss of $\pi^+\pi^-$ events.

Figure 13 shows the f_{loss} as a function of $M^{\pi^+\pi^-}_{inv}$ in cross-section bins for all JP triggers. The f_{loss} is independent of the triggers. The loss is higher in the lower mass, which decreases in the higher mass region. f_{loss} varies from 40% at the lower mass to 30% at the higher mass regions.
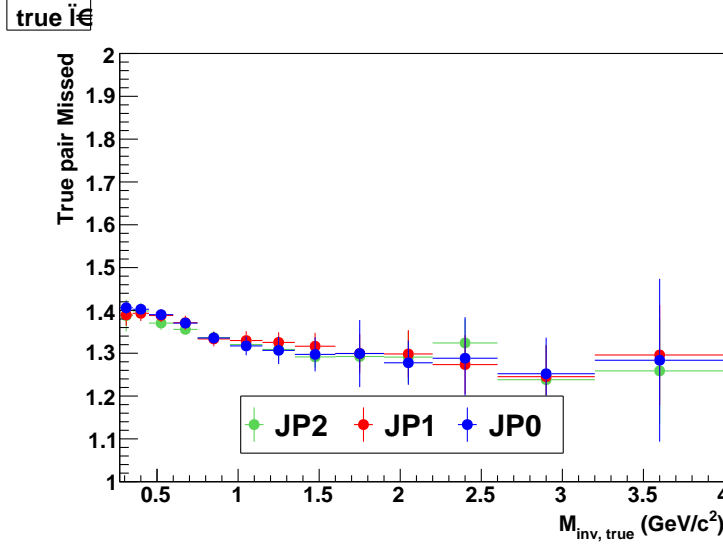
Figure 13: f_{loss} - PID cut efficiency fraction, which accounts for the loss of true $\pi^+\pi^-$ pairs due to the detector level PID cut, $-1 < n\sigma_\pi < 2$.

2.9 Measured Cross-Section Per Trigger $\left(\frac{d\sigma_{mes}^{\pi^+\pi^-}}{dM^{\pi^+\pi^-}}\right)$

With all components on hand as discussed above, the unpolarized $\pi^+\pi^-$ cross-section can be calculated using the formula;

$$\frac{d\sigma_{mes}^{\pi^+\pi^-}}{dM_{inv}^{\pi^+\pi^-}} = \frac{f_{fake} * f_{loss}}{L_{int} \cdot \epsilon_{trk}^{\pi^+\pi^-} \cdot \epsilon_{trg}^{\pi^+\pi^-}} \frac{dN_{true}^{\pi^+\pi^-}}{dM_{inv}^{\pi^+\pi^-}} \qquad (6)$$

Here, $dN_{true}^{\pi^+\pi^-}/dM_{inv}^{\pi^+\pi^-}$ is the unfolded yields normalized by the bin width. We use f_{fake} to correct for the fake pion contribution, as discussed in section 2.8.1, and f_{loss} to correct for the $n\sigma_\pi$ efficiency on the particle identification, as defined in section 2.8.2. L_{int} is the integrated luminosity in pb^{-1} (section 2.5). $\epsilon_{trk}^{\pi^+\pi^-}$ is the tracking efficiency correction factor, as discussed in section 2.7.1, and $\epsilon_{trg}^{\pi^+\pi^-}$ is the trigger efficiency correction factor, as discussed in section 2.7.2.

We measure the cross-section for each trigger and obtain the final cross-section by calculating the weighted average of all three triggered cross-sections. The weights for combining the cross-sections are determined from the statistical uncertainties in the final cross-section for each trigger.

$$w_i = \frac{1}{\delta_{stat,i}^2}, i = JP0, JP1, JP2 \qquad (7)$$

$$d\sigma_{mes}^{\pi^+\pi^-} = \frac{\sum_i w_i * d\sigma_{mes,i}^{\pi^+\pi^-}}{\sum_i w_i} \qquad (8)$$

$$\delta_{stat}^{\pi^+\pi^-} = \frac{1}{\sqrt{\sum w_i}} \qquad (9)$$

Figure 14 displays the measured cross-sections for all three triggers and the weighted average of all three triggered cross-sections (top panel), and the relative difference between the triggered and total cross-section in the bottom panel. We find good agreement between trigger-separated cross-sections. The relative difference between the triggered cross-sections and the final cross-section is
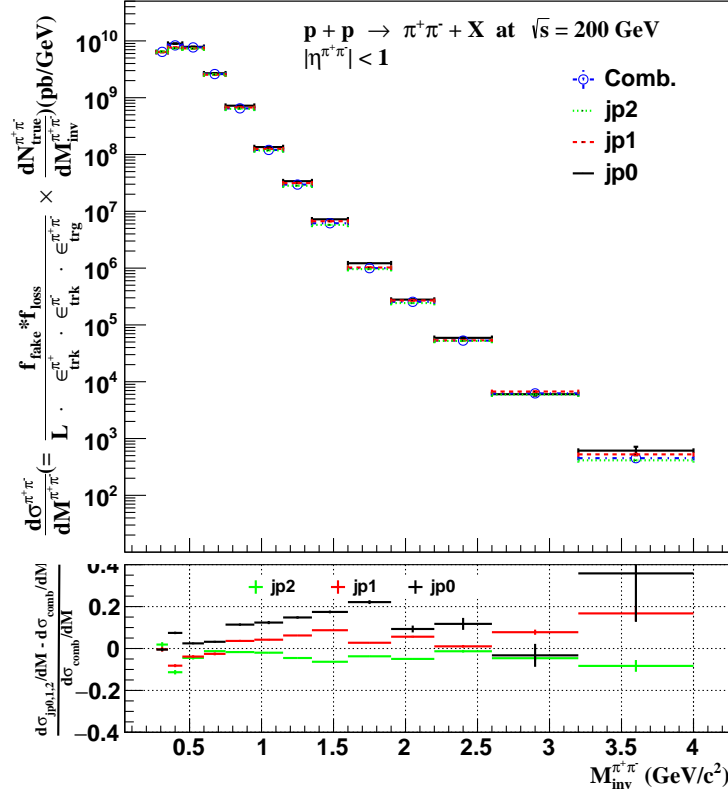
Figure 14: **Unpolarized $\pi^+\pi^-$ cross-sections per trigger** - Top panel: Unpolarized $\pi^+\pi^-$ cross-section as a function of $M_{inv}^{\pi^+\pi^-}$ for the JP0, JP1, and JP2 triggers. The final cross-section (labeled as "Comb") is the weighted average of all three triggered cross-sections. Bottom panel: the relative difference between triggered cross-sections and final cross-section.

applied as systematic uncertainty, which is discussed in section 2.7.2. Tables 7-9 in the appendix A present the numerical details for each triggered cross-section. The measured cross-section is reported in picobarns (pb).

2.10 Systematic Studies

The systematic uncertainties associated with this measurement are studied and quantified in the following sections:

2.10.1 Trigger Bias (δ_{bias})

While collecting data, a triggering algorithm is applied to select high transverse energy (E_T) jets suitable for physics analysis, which is otherwise shallowed by garbage low E_T jets. However, selecting only higher E_T jets biased the events favoring the quarks jets over gluon jets as gluon jets are broader, which may leak the energy to neighboring calorimeter towers, failing to meet the triggering condition. The size of this effect, called trigger bias, can be estimated using an embedding sample calculating the ratio of the quark fraction at the detector level to the same fraction at the particle level. Here, the detector level is where we apply the trigger to select hard $\pi^+\pi^-$ events, and the particle level is at the level before the trigger is applied. Mathematically, trigger bias fraction, f_{bias} can be expressed as:

$$f_{bias} = \frac{(N_q^{\pi^+\pi^-}/N_{q+g}^{\pi^+\pi^-})_{detector}}{(N_q^{\pi^+\pi^-}/N_{q+g}^{\pi^+\pi^-})_{particle}} \tag{10}$$

$$\delta_{bias} = 1 - f_{bias} \tag{11}$$

where, q and g referred to the quarks and gluons, respectively. The source parton, parton that is associated with the $\pi^+\pi^-$, is identified by matching the dipion to the parton by finding the parton that is closest in the $\eta - \phi$ space. The closest parton to the dipion in $\eta - \phi$ space is assigned as a source parton. Partons - quarks and gluons - associated with each $\pi^+\pi^-$ event are counted at the detector level and particle level. PDG codes are used to identify parton flavors and pions at the particle level.

The quark and gluon fraction calculated at the detector and particle level is shown in the left panel of figure 15. The quark fraction is shown in red, and the gluon fraction is shown in blue. The detector level fraction is represented by the dashed line, whereas the particle level fraction is represented by the solid line. The lower mass region is dominated by the gluon-initiated events, whereas the quark-initiated events dominate the higher mass region, as expected. The calculated fraction, f_{bias}, is shown in the right panel of figure 15. The associated systematic uncertainty, δ_{bias}, is given by equation 11. The δ_{bias} is found to be very small over the range of invariant mass bins.

2.10.2 Systematic uncertainty associated with the f_{fake} (δ_{fake})

The size of systematic uncertainty associated with the f_{fake} is estimated by varying the $n\sigma_\pi$ cut by 0.5 from both sides on the default PID cut $-1 < n\sigma_\pi < 2$. The difference between the ratio of the default fraction to the wider fraction ($-1.5 < n\sigma_\pi < 2.5$) and the ratio of the default fraction to the narrower fraction ($-0.5 < n\sigma_\pi < 1.5$) is assigned as a systematic uncertainty, δ_{fake}, as shown in figures 16-18 for each of the trigger. δ_{fake} is estimated for each of the cross-section bins.
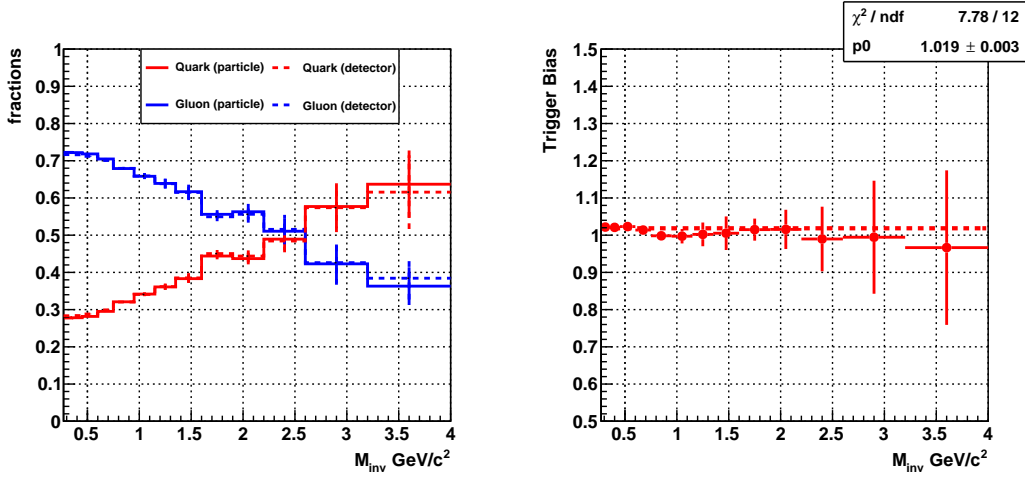
Figure 15: f_{bias} - Triggger bias fraction as a function of invariant mass.
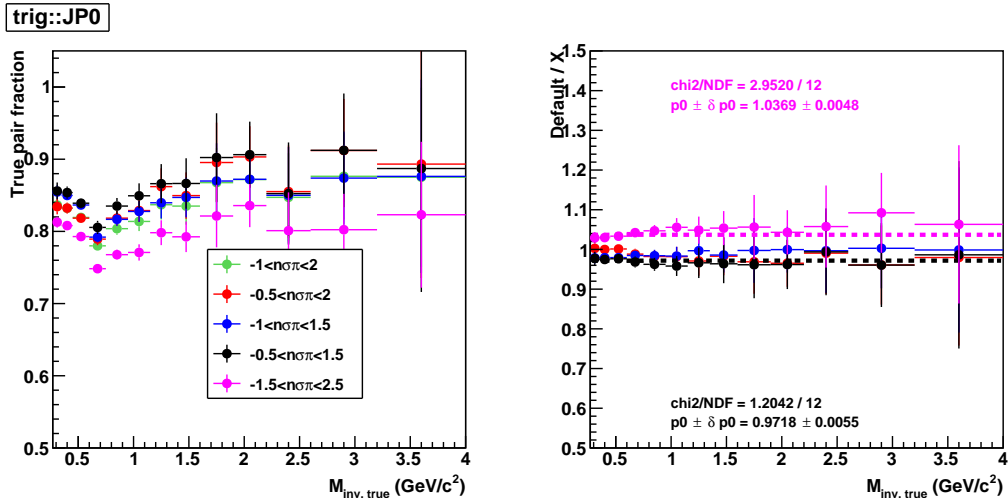
Figure 16: δ_{fake} **for JP0** - Systematic error associated to the f_{fake} for the JP0 trigger.
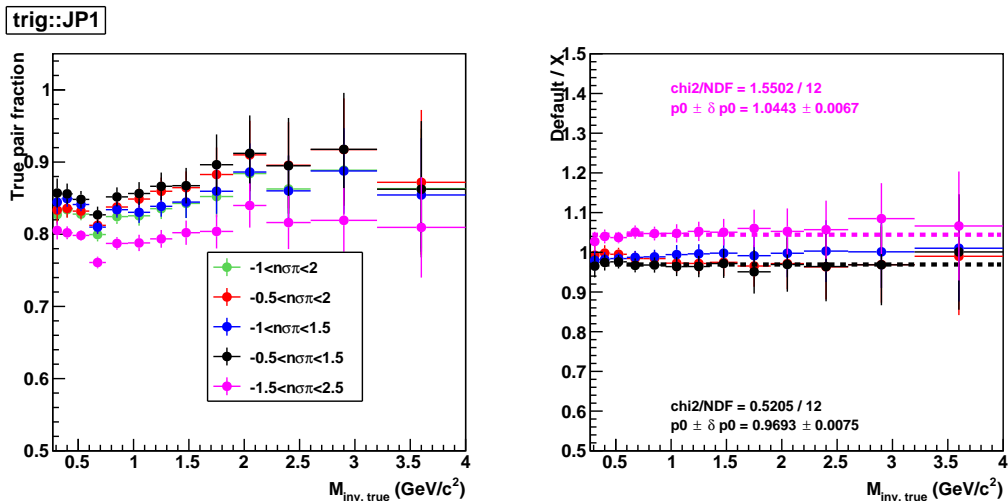
Figure 17: δ_{fake} **for JP1** - Systematic error associated to the f_{fake} for the JP1 trigger.
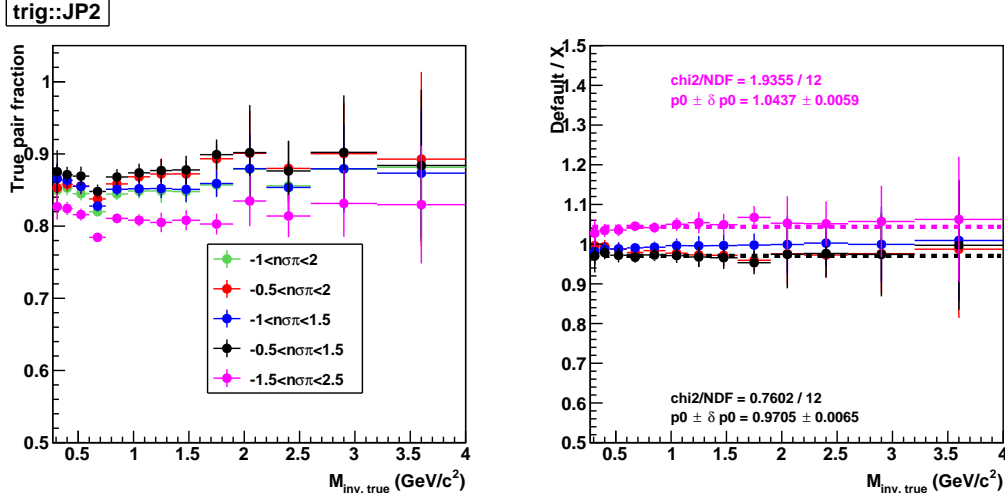
Figure 18: δ_{fake} **for JP2** - Systematic error associated to the f_{fake} for the JP2 trigger.

2.10.3 Systematic error associated with the f_{loss} (δ_{loss})

The systematic uncertainty associated with the f_{loss} is determined by studying the embedding description of the data. Figure 19 shows the reconstructed $n\sigma_{\pi}^{+}$ along the y-axis and $n\sigma_{\pi^-}$ along the x-axis of a true $\pi^+\pi^-$ pair from the embedding in the lowest $M_{inv}^{\pi^+\pi^-}$ bin for JP0 trigger. This is just a representative plot just to explain the general approach. All the data points displayed in the picture have a very good description of the data. However, embedding fails to describe the data in the region $n\sigma_{\pi} < -4$ (see figure 30), and the N_{true} fails to account for the $\pi^+\pi^-$ yields from this region.

To estimate the systematic uncertainty's size from f_{loss} (δ_{loss}), we examined the corresponding region on the opposite side ($n\sigma_{\pi} > 4$), where the embedding provides excellent data description. We expect similar trends in the $n\sigma_{\pi} < -4$ regions if the embedding accurately describes the entire data region. The δ_{loss} is calculated as the fraction of the true pair in the $|n\sigma_{\pi}| > 4$ regions,

$$\delta_{loss} = \frac{N_{true,|n\sigma_{\pi}|>4}}{N_{true}} \tag{12}$$

Figure 20 displays the δ_{loss} for JP0, JP1, and JP2 triggers. There is no trigger dependence in the δ_{loss}, which is about 6% in the overall mass region.

2.10.4 Cross-section discrepency between triggers (δ_{trg})

After correcting all of the triggered cross-sections, a discrepancy still remains, as shown in Figure 21, which displays the relative difference between the triggered and total cross-sections. This difference represents a systematic uncertainty associated with the trigger effect as a whole. Instead of using a weighted average, we calculate the maximum offset between any triggered cross-section and the final cross-section as our systematic uncertainty.

It might seem reasonable to expect this effect to be more pronounced in the lower mass bins and less so in the higher mass bins. However, we observe a larger difference in the higher mass bins, which we attribute to statistical fluctuations resulting from low statistics. To mitigate this, we combine the highest three bins for each trigger, as shown by the dashed lines in the figure. Then, we apply a single maximum offset value in those bins−in this case, the maximum offset is from the
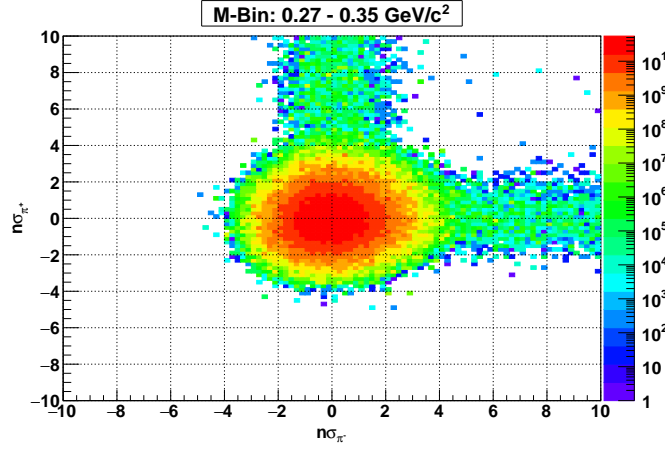
Figure 19: $n\sigma_{\pi^+}$ **vs** $n\sigma_{\pi^-}$ - Reconstructed $n\sigma_{\pi^+}$ vs $n\sigma_{\pi^-}$ of a true $\pi^+\pi^-$ pair in the lowest invariant mass bin.
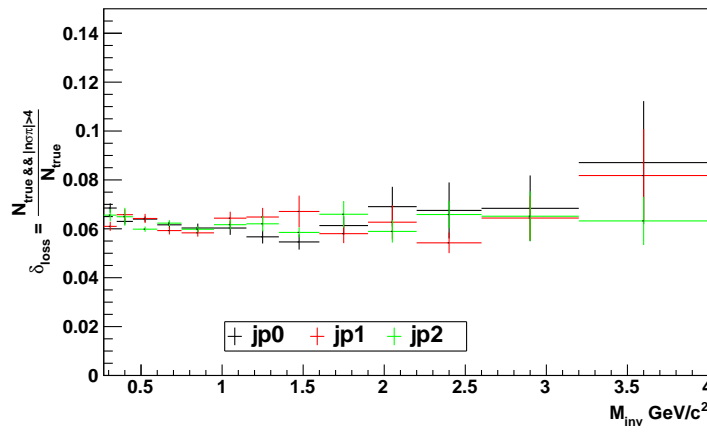
Figure 20: δ_{loss} - Systematic uncertainty associated with the f_{loss}.

₃₄₈ JP0−as a systematic uncertainty to the final cross-section.

$$\delta_{trg} = Max(\delta_{trg,JP0}, \delta_{trg,JP1}, \delta_{trg,JP2}) \tag{13}$$
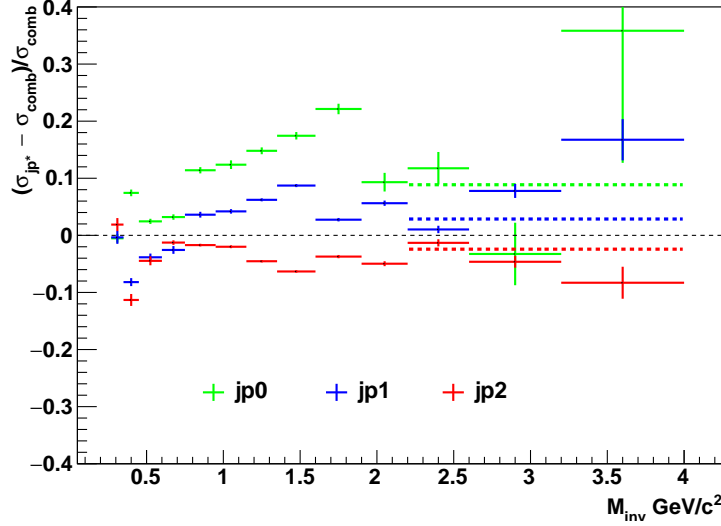
Figure 21: δ_{trg} - Trigger efficiency systematic uncertainty as a maximum relative difference between the triggered to the total cross-section. The last three bins are combined, and a single value of δ_{loss} is applied to those bins.

2.10.5 Embedding Sample Statistical Uncertainty ($\delta_{embstat}$)

₃₅₀ The statistical fluctuation due to the embedding sample statistics is calculated from the covariance
₃₅₁ of the response matrix used for the unfolding. The diagonal element (N_i) for each cross-section bin
₃₅₂ i of the covariance matrix gives the statistical uncertainty of the embedding sample:

$$\delta_{N,i} = \sqrt{N_i} \tag{14}$$

₃₅₃ The $\delta_{N,i}$ is statistical fluctuation at the unfolding level, which is propagated to the cross-section level
₃₅₄ by multiplying with the same correction factor ($f_{xsec} = \frac{f_{fakr}*f_{loss}}{L*BW*\epsilon_{trg}*\epsilon_{trk}}$) applied for the cross-section
₃₅₅ (equation 6):

$$\delta_{embstat,i} = f_{xsec,i} \times \delta_{N,i} \tag{15}$$

₃₅₆ The $\delta_{embstat,i}$ is calculated for each trigger used, by using the diagonal element of the covariance
₃₅₇ matrices, which are shown in figure 22. The weighted average of all $\delta_{embstat,i}$ values from all triggers
₃₅₈ is then applied as a systematic uncertainty due to the embedding sample statistics to the final
₃₅₉ cross-section result.

2.10.6 Total systematic uncertainties ($\delta\sigma_{sys}^{Syst.}$)

₃₆₁ The total systematic uncertainties for the final cross-section is the quadrature sum of all the sys-
₃₆₂ tematic uncertainties discussed above. The δ_{fake} and δ_{loss} are the weighted averages of all triggered
₃₆₃ cases. The δ_{bias} is calculated for the combined trigger, which directly applies to the final cross-
₃₆₄ section. The δ_{trg} is also directly applied to the final cross-section as it is calculated as the maximum
₃₆₅ discrepancy between the triggered and the final cross-section.

25
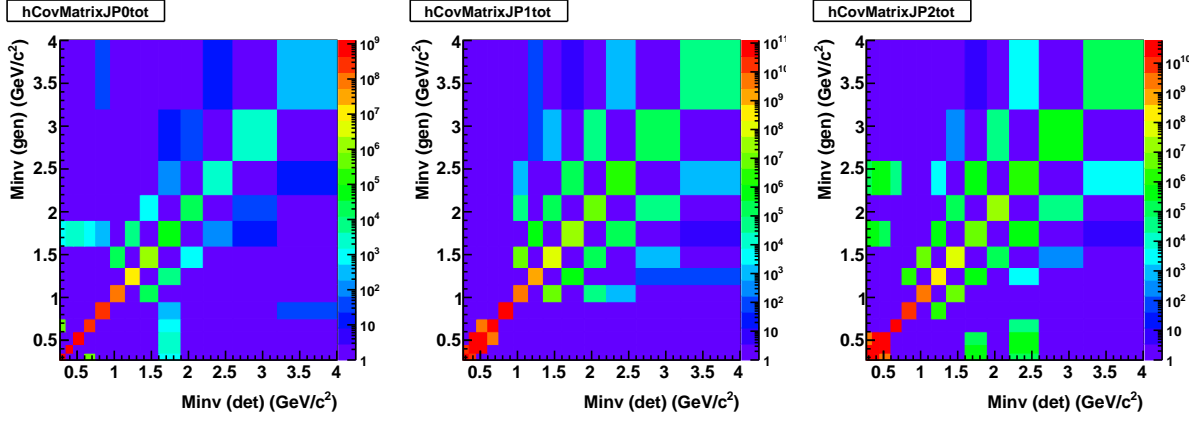
Figure 22: **Covariance Matrices** - Covariance matrices of the response matrix for JP0, JP1, and JP2 triggers from left to right, respectively.

$$\delta\sigma_{mes}^{Syst.} = \sqrt{\delta_{trg}^2 + \delta_{fake}^2 + \delta_{loss}^2 + \delta_{bias}^2 + \delta_{embstat}^2} \tag{16}$$

Figure 23 displays a breakdown of each of the contributing systematic uncertainties to the final cross-section. The dominant systematic uncertainty comes from the trigger effect at the final cross-section. Numerical values of all the systematic uncertainties are presented in the table 4.

Bin No.	δ_{trg}	δ_{fake}	δ_{loss}	δ_{bias}	$\delta_{embstat}$	$\delta\sigma_{mes}^{Syst.}$
0	0.0189	0.0343	0.067	0.0218	0.00897	0.081
1	0.113	0.0392	0.0639	0.021	0.00626	0.138
2	0.0447	0.039	0.064	0.0231	0.00509	0.0904
3	0.032	0.0627	0.0613	0.0138	0.00682	0.0946
4	0.114	0.0789	0.0601	0.00169	0.0107	0.152
5	0.124	0.0992	0.0612	0.00281	0.012	0.17
6	0.148	0.108	0.0594	0.00214	0.0121	0.193
7	0.174	0.1	0.0589	0.00526	0.0104	0.21
8	0.221	0.0907	0.0599	0.0149	0.00896	0.247
9	0.0931	0.0841	0.0663	0.0155	0.0224	0.144
10	0.0131	0.0952	0.0613	0.0102	0.0244	0.117
11	0.0778	0.0969	0.0669	0.00551	0.0452	0.148
12	0.167	0.0573	0.0854	0.0334	0.12	0.232

Table 4: Final systematic uncertainties.
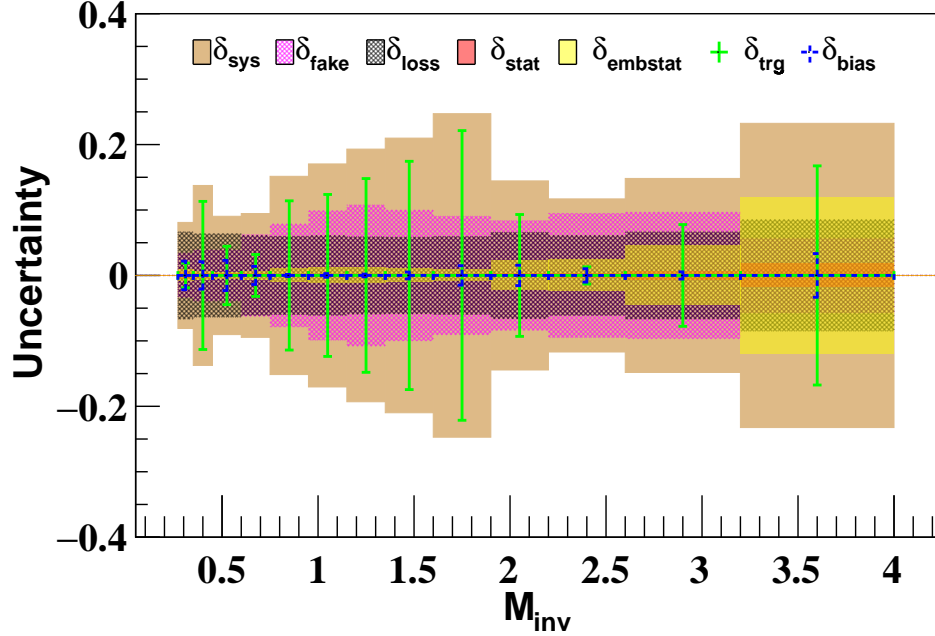
Figure 23: δ_{sys} - Breakdown of systematic uncertainties contributing to the final cross-section result.

3 Final cross-section result

3.1 PYTHIA cross-section

The PYTHIA cross-section is calculated using the standalone PYTHIA events from the embedding sample used in this analysis. All PYTHIA events are used, and the fiducial kinematic cuts on track p_T and η, as applied in the measured cross-section (see table 1), are retained at the PYTHIA level as well. The PYTHIA pions are selected using pdg codes (211 for π^+ and -211 for π^-). The $\pi^+\pi^-$ construction procedure is the same as in the measured construction, as discussed in section 2.3.

The PYTHIA unpolarized $\pi^+\pi^-$ cross-section is calculated for each of the partonic p_T bins. The embedding sample is produced in the 13 partonic p_T ranges from 2 GeV to infinity. The total number of $\pi^+\pi^-$ events ($N_{p_T}^{\pi^+\pi^-}$) are counted and binned as a function of $M_{inv}^{\pi^+\pi^-}$ for each partonic p_T bins. The PYTHIA cross-section binning is the same as in the measured cross-section. The number of collision events processed ($N_{p_T}^{pro}$), and the cross-section (σ_{p_T}) is used to calculate luminosity (L_{p_T}) for each partonic p_T bin as,

$$L_{p_T} = \frac{N_{p_T}^{pro}}{\sigma_{p_T}} \tag{17}$$

The unpolarized $\pi^+\pi^-$ cross-section for each p_T bin is then calculated as,

$$\frac{d\sigma_{p_T}^{\pi^+\pi^-}}{dM^{\pi^+\pi^-}} = \frac{N_{p_T}^{\pi^+\pi^-}}{L_{p_T} * BinWidth} \tag{18}$$

p_T Bin	$N_{p_T}^{tried}$	$N_{p_T}^{gen}$	$N_{p_T}^{pro}$	σ_{p_T}	$\delta\sigma_{p_T}$	L (pb^{-1})	δL (pb^{-1})
2−3	1.77e+07	3.69e+06	3.33e+06	9e+09	1.9e+08	0.00037	7.81e-06
3−4	1.43e+07	3.69e+06	3.31e+06	1.46e+09	3.39e+07	0.00226	5.25e-05
4−5	1.52e+07	3.69e+06	3.3e+06	3.54e+08	8.68e+06	0.0093	0.000228
5−7	1.79e+07	3.69e+06	3.28e+06	1.52e+08	3.54e+06	0.0217	0.000506
7−9	1.91e+07	3.69e+06	3.29e+06	2.49e+07	5.76e+05	0.132	0.00305
9−11	1.94e+07	3.69e+06	3.28e+06	5.85e+06	1.37e+05	0.561	0.0131
11−15	2.2e+07	3.69e+06	3.28e+06	2.3e+06	5.51e+04	1.42	0.0341
15−20	2.52e+07	3.69e+06	3.28e+06	3.43e+05	8.14e+03	9.56	0.227
20−25	3.03e+07	3.69e+06	3.27e+06	4.57e+04	1.09e+03	71.6	1.71
25−35	2.26e+07	2.46e+06	2.18e+06	9.73e+03	275	224	6.34
35−45	2.35e+07	2.46e+06	2.18e+06	470	12.1	4.65e+03	120
45−55	1.27e+07	1.23e+06	1.09e+06	26.9	0.709	4.06e+04	1.07e+03
55−−1	1.25e+07	1.23e+06	1.09e+06	1.43	0.0398	7.61e+05	2.11e+04

Table 5: PYTHIA number of events and luminosity per partonic p_T bins. Cross-section (σ_{p_T}) is in picobarn (pb), luminosity (L) is in pb^{-1}, and partonic p_T is in GeV/c.

The final cross-section is the sum of all cross-sections from each of the partonic p_T bin, which is shown (in black) in figure 14 (top panel), which follows a similar trend to the measured cross-section. The agreement of PYTHIA with the measured cross-section is within 10% up to $M_{inv}^{\pi^+\pi^-} \sim 2.25$ GeV/c^2, above which the disagreement grows up to $\sim 50\%$.

3.2 Measured cross-section

The measured cross-section is the weighted average of JP0, JP1, and JP2 triggered cross-sections, shown in red in the top panel of figure 24. The absolute PYTHIA cross-section is also included (shown in black) for comparison. The bottom panel shows the final systematic uncertainty (green band), the statistical uncertainty (red band), and the relative difference between the PYTHIA and the measured cross-section (red lines) relative to the measured cross-section. The 10% uncertainty from the luminosity measurement is not included in the final uncertainty.
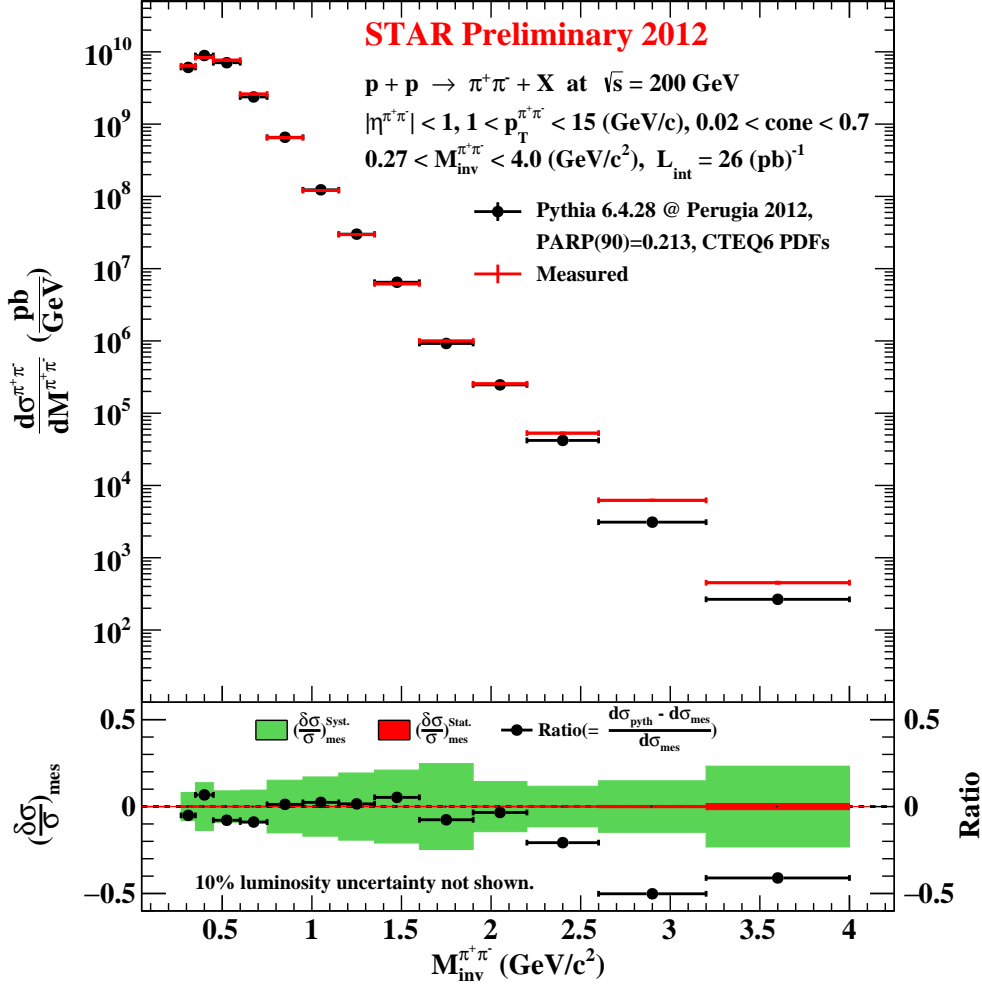
Figure 24: **Unpolarized $\pi^+\pi^-$ cross-section** - Top panel: measured unpolarized $\pi^+\pi^-$ cross-section and PYTHIA cross-section as a function of the $M_{inv}^{\pi^+\pi^-}$. Bottom panel: Relative difference between the measured and the PYTHIA cross-section on the right y-axis (Ratio $= \frac{d\sigma_{pyth}/dM - d\sigma_{mes}/dM}{d\sigma_{mes}/dM}$), and relative statistical ($(\frac{\delta\sigma}{\sigma})_{mes}^{Stat.}$) and systematic ($(\frac{\delta\sigma}{\sigma})_{mes}^{Syst.}$) errors on the left y-axis.

$M(GeV/c^2)$	$d\sigma_{mes}^{\pi^+\pi^-}$ (pb)	$\delta\sigma_{mes}^{Stat.}$ (pb)	$d\sigma_{pyth}^{\pi^+\pi^-}$ (pb)	$\delta\sigma_{pyth}^{Stat.}$ (pb)	$(\delta\sigma/d\sigma)_{mes}^{Syst.}$
0.27-0.35	6.41e+09	2.99e+07	6.09e+09	9.54e+07	0.081
0.35-0.45	8.33e+09	2.89e+07	8.89e+09	1.38e+08	0.138
0.45-0.6	7.68e+09	2.15e+07	7.07e+09	1.08e+08	0.0904
0.6-0.75	2.61e+09	6.13e+06	2.38e+09	3.41e+07	0.0946
0.75-0.95	6.49e+08	1.01e+06	6.57e+08	8.72e+06	0.152
0.95-1.15	1.2e+08	1.96e+05	1.23e+08	1.65e+06	0.17
1.15-1.35	2.96e+07	3.86e+04	3e+07	4.86e+05	0.193
1.35-1.6	6.16e+06	8.52e+03	6.48e+06	1.41e+05	0.21
1.6-1.9	9.98e+05	1.66e+03	9.22e+05	1.43e+04	0.247
1.9-2.2	2.55e+05	648	2.46e+05	5.02e+03	0.144
2.2-2.6	5.3e+04	194	4.2e+04	1.33e+03	0.117
2.6-3.2	6.23e+03	39.7	3.11e+03	144	0.148
3.2-4	451	8	266	16	0.232

Table 6: Final cross-section results.

4 Summary

We conducted a novel experimental measurement of the unpolarized $\pi^+\pi^-$ cross-section as a function of $\pi^+\pi^-$ invariant mass using the STAR Run 12 proton-proton collision dataset at $\sqrt{s} = 200$ GeV. This marks the first-ever experimental measurement of its kind. Our findings serve to constrain the unpolarized fragmentation function, and in particular, the gluon fragmentation function. Additionally, the high-precision IFF asymmetries obtained from Run 15, combined with this measurement, will provide a foundation for transversity extraction with greater precision than previously possible.

5 Outlook

Our upcoming publication will report on the results of this measurement and the 2015 IFF asymmetry measurement. However, before publication, we need to address a few details to ensure the maximum impact of our results. Currently, our measured cross section is only in the invariant mass bins, but we recognize that our findings would have been more impactful if measured in both the mass and p_T bins using 2D unfolding techniques. Additionally, we need to optimize the cone cut (max and min) to ensure that the pairs originate from the same parton or within the same jet, as required. Moreover, at the preliminary level, correction factors associated with the particle identification are studied based on the embedding sample, provided that data and embedding have an excellent agreement. This will be further explored with the standard data-driven approach, providing the consistency test between the two methods. We will thoroughly study all these components to ensure the final results are carefully evaluated.

References

[1] J. Ashman et al. A measurement of the spin asymmetry and determination of the structure function g1 in deep inelastic muon-proton scattering. *Phys. Lett. B*, 206:364, 1988.

[2] L. Adamczyk et al. Observation of Transverse Spin-Dependent Azimuthal Correlations of Charged Pion Pairs in $p^\uparrow + p$ at $\sqrt{s} = 200$ GeV. *Phys. Rev. Lett.*, 115:242501, 2015.

[3] L. Adamczyk et al. Transverse spin-dependent azimuthal correlations of charged pion pairs measured in p$^\uparrow$+p collisions at $\sqrt{s} = 500$ GeV. *Phys. Lett. B*, 780:332–339, 2018.

[4] Babu R. Pokhrel. Measurement of Transverse Spin Dependent Azimuthal Correlations of Charged Pion(s) in $p^\uparrow p$ Collisions at $\sqrt{s} = 200$ GeV at STAR. *SciPost Phys. Proc.*, 8:047, 2022.

[5] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

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

[7] James Kevin Adkins. *Studying Transverse Momentum Dependent Distributions in Polarized Proton Collisions via Azimuthal Single Spin Asymmetries of Charged Pions in Jets.* PhD thesis, Kentucky U., 2015.

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

[9] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini. GEANT3. 9 1987.

[10] Dmitrii Kalinkin. *Measurement of Mid-rapidity Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 200$ GeV. PhD thesis, Indiana University, August* 2022.

[11] Stefan Schmitt. TUnfold: an algorithm for correcting migration effects in high energy physics. JINST, 7:T10003, 2012.

A Triggerd Cross-Section Results

A.1 JP0 cross-section

Bin	BW	L	ϵ_{trg}	ϵ_{trk}	f_{fake}	f_{loss}	$N_{un,true}$	$\delta_{un,stat}$	$\sigma^{\pi^+\pi^-}$	$\delta\sigma^{\pi^+\pi^-}_{stat}$
0	0.08	0.16	0.0622	0.888	0.839	1.41	3.82e+06	2.33e+04	6.38e+09	3.89e+07
1	0.1	0.16	0.0636	0.891	0.836	1.4	6.92e+06	3.02e+04	8.95e+09	3.91e+07
2	0.15	0.16	0.0666	0.869	0.815	1.39	9.65e+06	3.33e+04	7.87e+09	2.72e+07
3	0.15	0.16	0.0843	0.881	0.742	1.37	4.71e+06	1.89e+04	2.69e+09	1.08e+07
4	0.2	0.16	0.115	0.889	0.779	1.34	2.28e+06	1.08e+04	7.23e+08	3.42e+06
5	0.2	0.16	0.163	0.889	0.784	1.32	6.07e+05	3.99e+03	1.35e+08	8.9e+05
6	0.2	0.16	0.228	0.895	0.813	1.31	2.08e+05	1.06e+03	3.39e+07	1.72e+05
7	0.25	0.16	0.319	0.906	0.858	1.3	7.51e+04	407	7.23e+06	3.92e+04
8	0.3	0.16	0.5	0.91	0.877	1.3	2.33e+04	173	1.22e+06	9.02e+03
9	0.3	0.16	0.558	0.935	0.901	1.28	6.08e+03	89.6	2.79e+05	4.11e+03
10	0.4	0.16	0.639	0.914	0.889	1.29	1.93e+03	49	5.92e+04	1.5e+03
11	0.6	0.16	0.761	0.92	0.863	1.25	375	21.1	6.03e+03	339
12	0.8	0.16	0.901	0.888	0.847	1.27	58.1	9.87	612	104

Table 7: Trigger JP0 cross section measurement details by numbers. Luminosity (L) is in pb^{-1} and cross-section is in pb.

A.2 JP1 cross-section

Bin	BW	L	ϵ_{trg}	ϵ_{trk}	f_{fake}	f_{loss}	$N_{un,true}$	$\delta_{un,stat}$	$\sigma^{\pi^+\pi^-}$	$\delta\sigma^{\pi^+\pi^-}_{stat}$
0	0.08	7.68	0.00731	0.885	0.833	1.39	2.2e+07	2.27e+05	6.39e+09	6.6e+07
1	0.1	7.68	0.00774	0.885	0.832	1.39	3.47e+07	2.29e+05	7.65e+09	5.06e+07
2	0.15	7.68	0.0083	0.862	0.822	1.39	5.33e+07	3.27e+05	7.38e+09	4.53e+07
3	0.15	7.68	0.0123	0.883	0.765	1.37	3.04e+07	1.99e+05	2.54e+09	1.67e+07
4	0.2	7.68	0.0207	0.891	0.808	1.33	1.77e+07	8.24e+04	6.73e+08	3.13e+06
5	0.2	7.68	0.0372	0.897	0.811	1.33	5.97e+06	2.03e+04	1.25e+08	4.27e+05
6	0.2	7.68	0.0601	0.9	0.826	1.33	2.37e+06	5.07e+03	3.14e+07	6.7e+04
7	0.25	7.68	0.104	0.901	0.858	1.32	1.06e+06	2.3e+03	6.7e+06	1.45e+04
8	0.3	7.68	0.212	0.92	0.868	1.3	4.08e+05	997	1.02e+06	2.5e+03
9	0.3	7.68	0.257	0.925	0.887	1.3	1.28e+05	468	2.69e+05	982
10	0.4	7.68	0.347	0.927	0.885	1.27	4.71e+04	250	5.35e+04	284
11	0.6	7.68	0.507	0.922	0.867	1.25	1.34e+04	130	6.72e+03	65.2
12	0.8	7.68	0.642	0.871	0.813	1.28	1.73e+03	46.5	526	14.1

Table 8: Trigger JP1 cross section measurement details by numbers. Luminosity (L) is in pb^{-1} and cross-section is in pb.

A.3 JP2 cross-section

Bin	BW	L	ϵ_{trg}	ϵ_{trk}	f_{fake}	f_{loss}	$N_{un,true}$	$\delta_{un,stat}$	$\sigma^{\pi^+\pi^-}$	$\delta\sigma^{\pi^+\pi^-}_{stat}$
0	0.08	18.8	0.00139	0.876	0.845	1.39	1.02e+07	1.04e+05	6.53e+09	6.68e+07
1	0.1	18.8	0.00151	0.882	0.843	1.4	1.57e+07	1.75e+05	7.39e+09	8.23e+07
2	0.15	18.8	0.00169	0.854	0.838	1.37	2.6e+07	1.97e+05	7.34e+09	5.56e+07
3	0.15	18.8	0.00277	0.88	0.806	1.36	1.62e+07	5.24e+04	2.57e+09	8.33e+06
4	0.2	18.8	0.00541	0.884	0.833	1.34	1.03e+07	1.82e+04	6.38e+08	1.13e+06
5	0.2	18.8	0.0109	0.893	0.838	1.32	3.89e+06	7.5e+03	1.18e+08	2.27e+05
6	0.2	18.8	0.0206	0.901	0.847	1.31	1.77e+06	3.08e+03	2.82e+07	4.91e+04
7	0.25	18.8	0.0397	0.909	0.86	1.29	8.81e+05	1.67e+03	5.77e+06	1.09e+04
8	0.3	18.8	0.0874	0.913	0.861	1.29	3.88e+05	926	9.6e+05	2.29e+03
9	0.3	18.8	0.122	0.911	0.876	1.29	1.34e+05	487	2.42e+05	882
10	0.4	18.8	0.174	0.932	0.875	1.32	5.53e+04	286	5.23e+04	270
11	0.6	18.8	0.301	0.918	0.879	1.24	1.7e+04	145	5.94e+03	50.7
12	0.8	18.8	0.439	0.862	0.846	1.26	2.21e+03	52.1	413	9.75

Table 9: Trigger JP2 cross-section measurement details by numbers. Luminosity (L) is in pb^{-1} and cross-section is in pb.

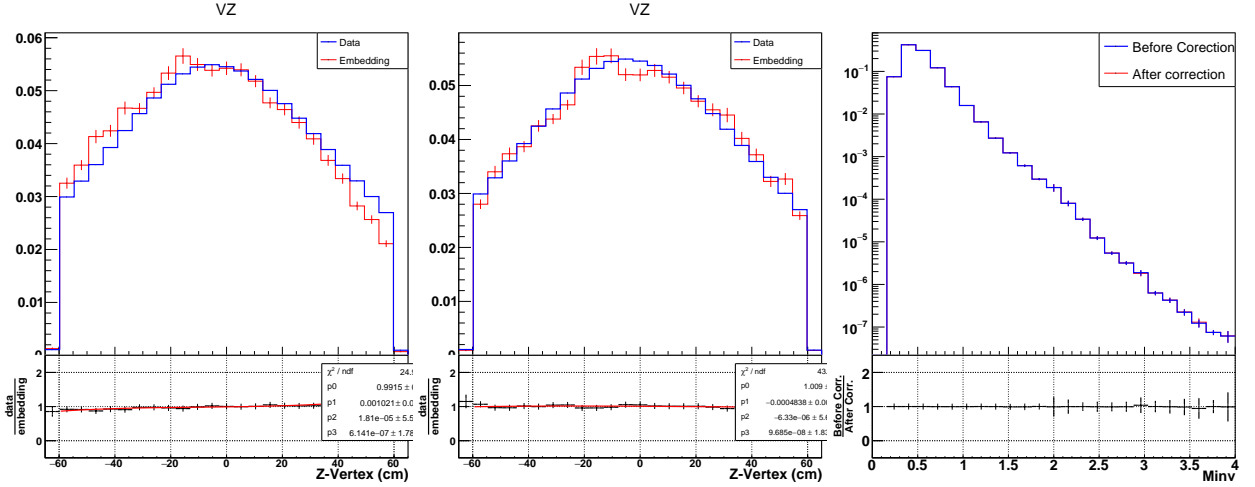# B Data Embedding Comparison

B.1 Vertex Shape Correction

Figure 25: Vertex shift (*left*), after vertex correction (*middle*), and impact of vertex correction in invariant mass distribution for JP0 trigger.

For the embedding sample to be matched with the data, the event z-vertices should be matched. Once z-vertex matches, other distributions should also be matched. Figure 25(*left*) shows the z-vertices comparison between data and embedding, where embedding vertex distribution is shifted to the left. This mismatch cannot be ignored and needs correction. It is easy to fix this type of

mismatch in physics analysis by taking the ratio between the two distributions (*bottom panel*) and fitting it with the best polynomial fit. The fit equation, with the z-vertex as a variable, can then be applied as a weight to the z-vertex, which should fix the mismatch. The middle plot shows the comparison after the z-vertex shape correction. The impact of this shape correction is checked by comparing the exact same distribution before and after the correction and finding that there is NO impact of it on the rest of the distributions, as shown in the right plot.

In this analysis, only JP0, JP1, and JP2 events are used, whereas in the data, the events are heavily prescaled while production. This means that only one event is kept for every event (*prescale*). In data, the triggering algorithm works in such a way that it first checks the highest threshold trigger, for example, JP2, followed by lower threshold triggers, for example, JP1, and JP0, in that order. If an event meets JP2 criteria, it will meet all the lower trigger categories. This may occur trigger overlapping and have a chance that the same event would be counted for each trigger. To overcome trigger overlapping issues in the data, prescale factor is applied at the production level. Prescale for the JP2 is always one, and the prescale value is larger for the lower threshold triggers.

However, the embedding sample takes everything, as there is no triggering scheme in the simulation. Trigger information can be accessed using the trigger simulator in the embedding sample. To replicate the triggering scheme as in the data and overcome trigger overlapping, the same prescale factors from the data should be applied for each trigger. Individual trigger event distributions can be directly compared with the data. However, one should be careful while combining all triggers in the embedding. For combined triggers, the correlation between data and embedding can be established as follows:

$$H_{d,JP0} + H_{d,JP1} + H_{d,JP2} \leftrightarrow \frac{1}{p_{JP0}}H_{e,JP0} + \frac{1}{p_{JP1}}H_{e,JP0} + \frac{1}{p_{JP2}}H_{e,JP2} \tag{19}$$

where, H is distributions from embedding(e) and data(d), and p is prescale factor for each individual triggers JP0, JP1, and JP2. The following section includes data-embedding comparison plots for all three triggers and also for combined triggers. Embedding distributions match well with the data.
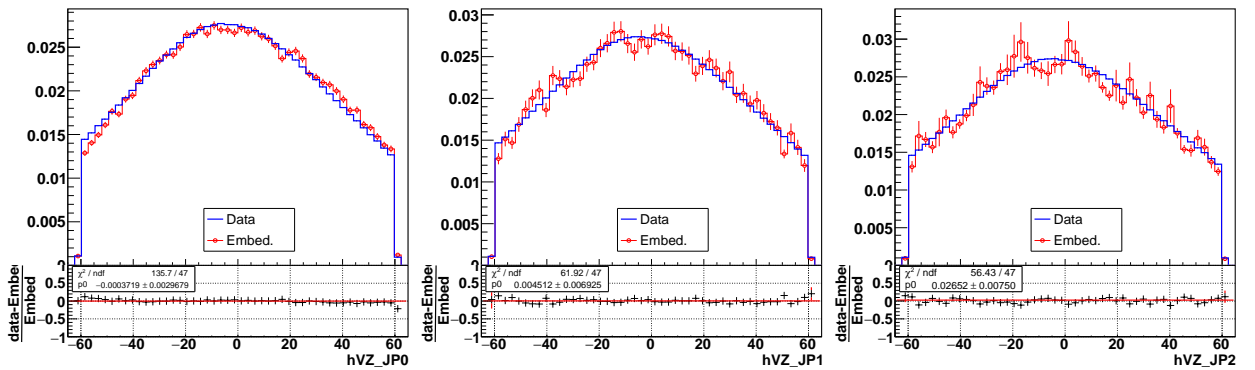
B.2   Data-Embedding Comparison plots

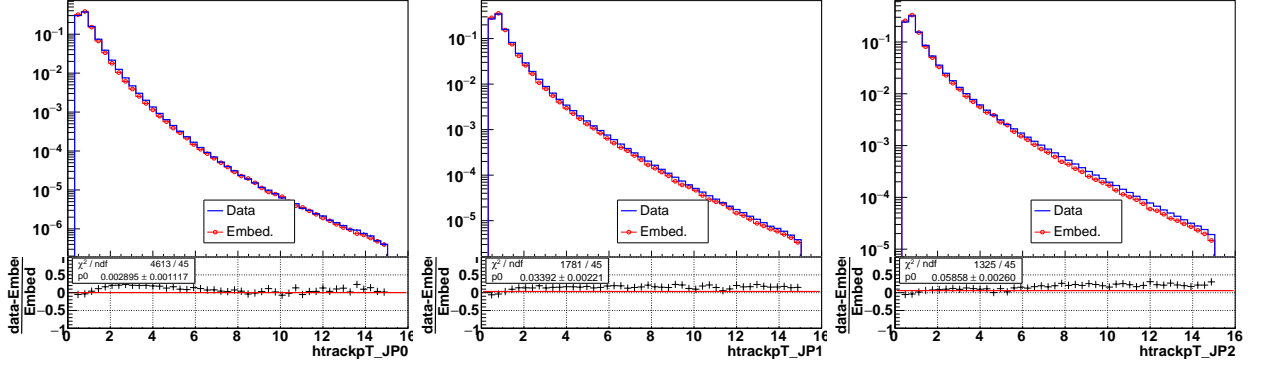Figure 26: Z-vertex distribution comparison for JP0, JP1, and JP2 from left to right respectively.

Figure 27: Hadrons p_T distribution comparison for JP0, JP1, and JP2 from left to right respectively.

Figure 28: Hadrons η distribution comparison for JP0, JP1, and JP2 from left to right respectively.

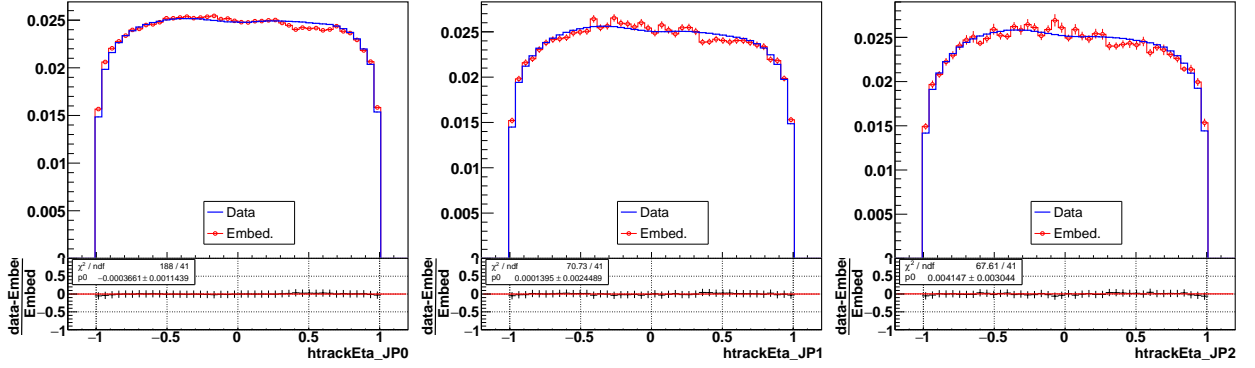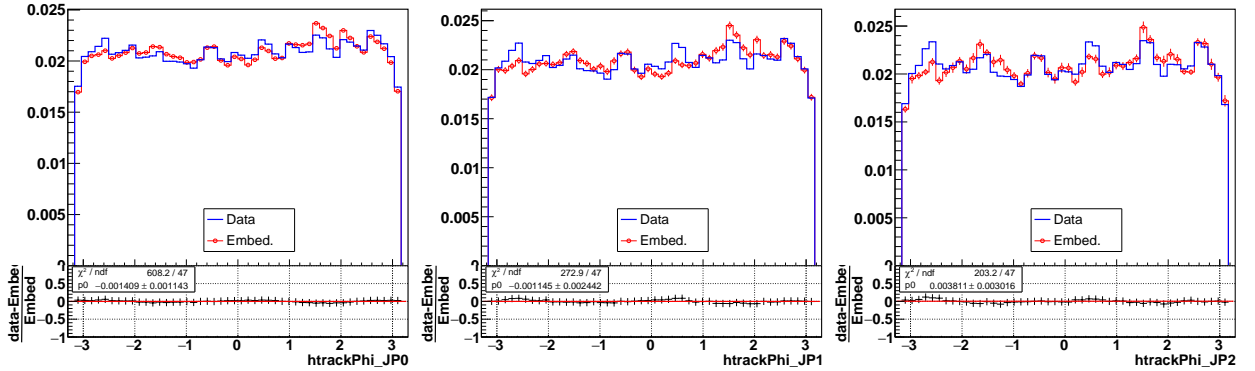Figure 29: Hadrons ϕ distribution comparison for JP0, JP1, and JP2 from left to right respectively.
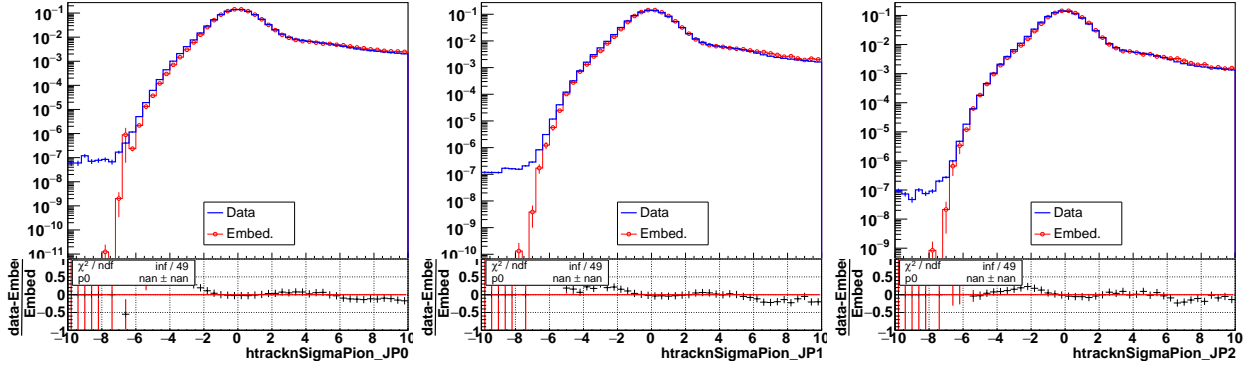
Figure 30: Hadrons $n\sigma\pi$ distribution comparison for JP0, JP1, and JP2 from left to right respectively.
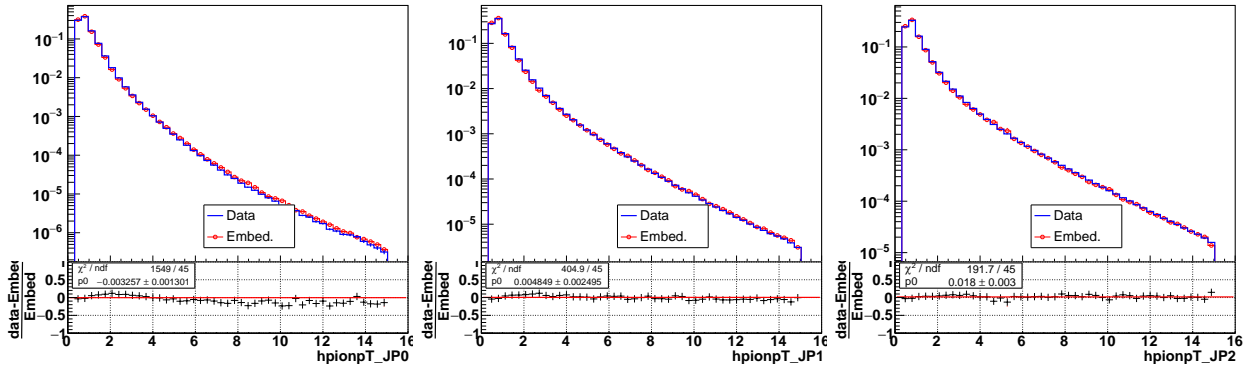

Figure 31: Pion p_T distribution comparison for JP0, JP1, and JP2 from left to right respectively.

Figure 32: Pion ϕ distribution comparison for JP0, JP1, and JP2 from left to right respectively.

Figure 33: Pion η distribution comparison for JP0, JP1, and JP2 from left to right respectively.

Figure 34: $\pi^+\pi^-$ opening angle (in $\eta-\phi$ space) distribution comparison for JP0, JP1, and JP2 from left to right respectively.

Figure 35: $\pi^+\pi^-$ p_T distribution comparison for JP0, JP1, and JP2 from left to right respectively.
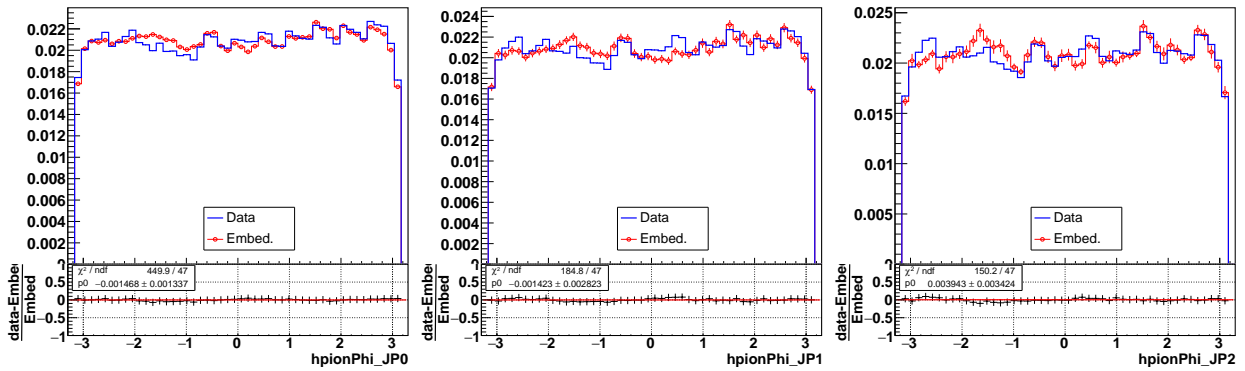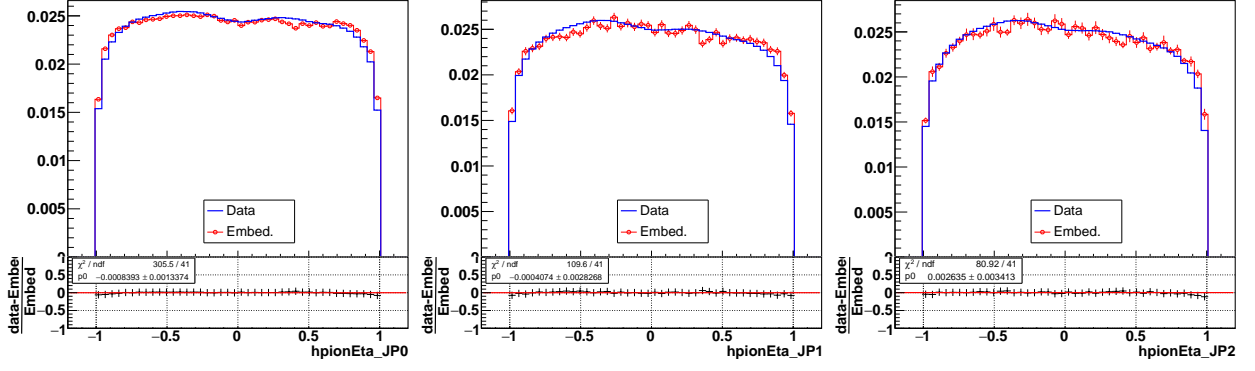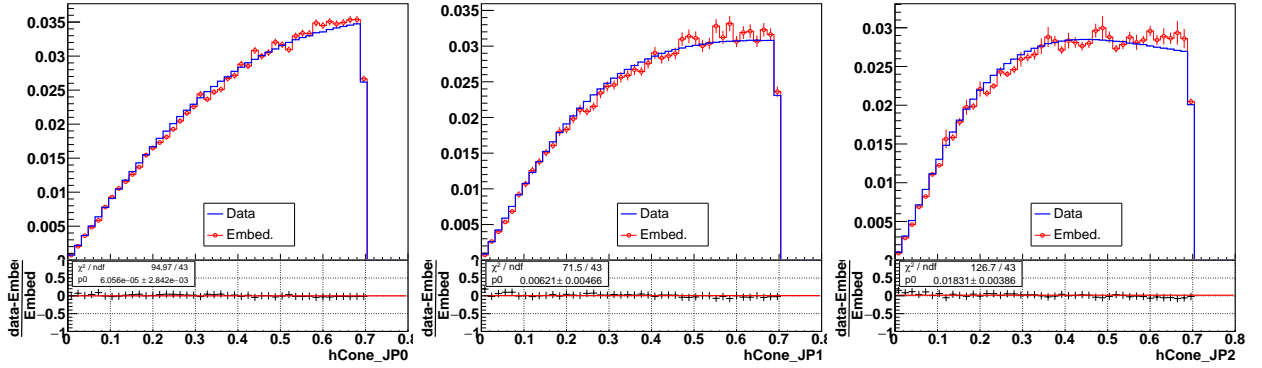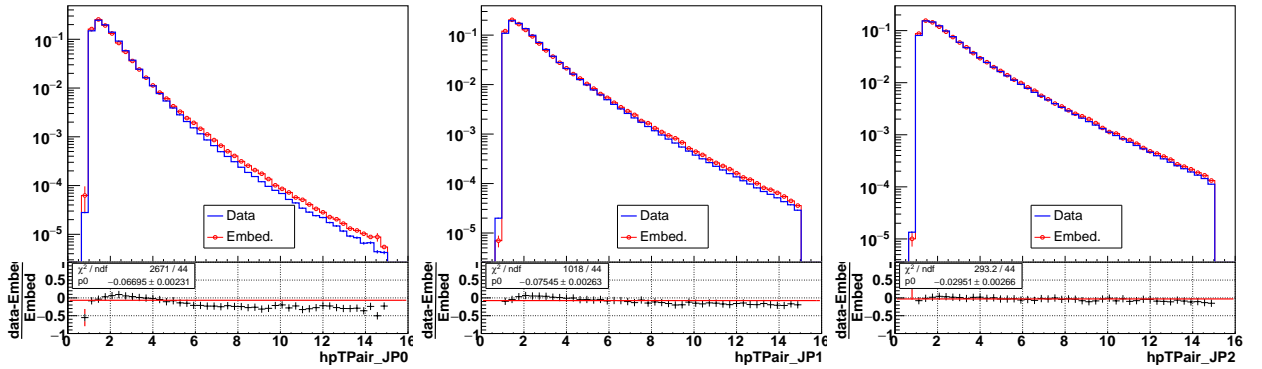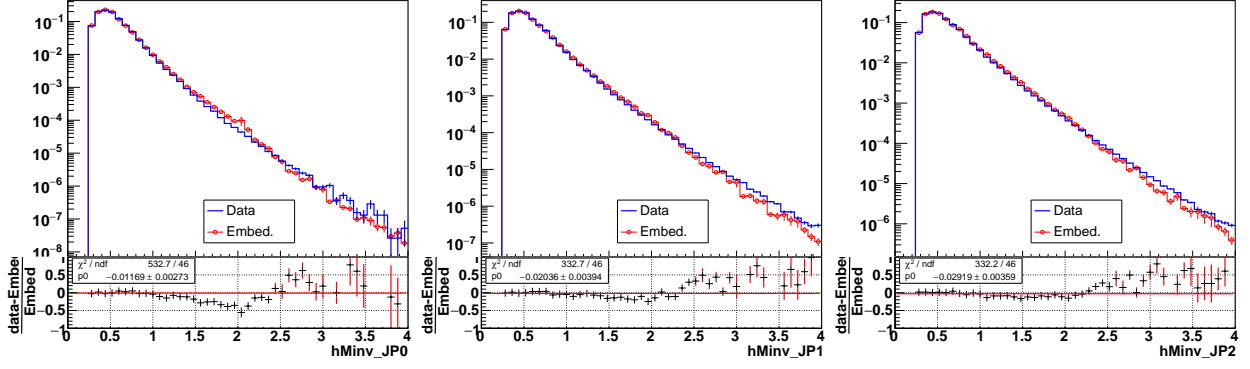
Figure 36: $M_{inv}^{\pi^+\pi^-}$ distribution comparison for JP0, JP1, and JP2 from left to right respectively.
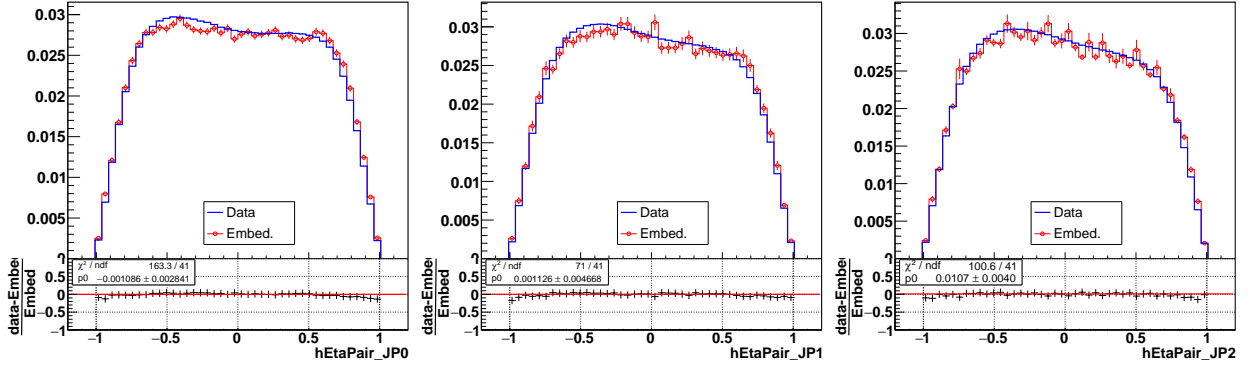

Figure 37: $\eta^{\pi^+\pi^-}$ distribution comparison for JP0, JP1, and JP2 from left to right respectively.

C   Run 2012 List (Data)

13047003 13047004 13047022 13047023 13047024 13047026 13047027 13047028 13047029 13047030
13047031 13047032 13047033 13047034 13047035 13047036 13047037 13047039 13047041 13047042
13047043 13047044 13047122 13047123 13047124 13047126 13048009 13048010 13048011 13048012
13048013 13048014 13048015 13048016 13048017 13048018 13048019 13048030 13048031 13048032
13048040 13048041 13048042 13048043 13048044 13048045 13048046 13048049 13048050 13048051
13048052 13048053 13048087 13048088 13048089 13048090 13048091 13048092 13048093 13049004
13049005 13049006 13049007 13049031 13049032 13049035 13049039 13049041 13049042 13049044
13049045 13049046 13049047 13049048 13049049 13049050 13049072 13049073 13049080 13049081
13049082 13049086 13049087 13049088 13049089 13049093 13049094 13049096 13049098 13049099
13049101 13050001 13050006 13050007 13050009 13050011 13050012 13050015 13050016 13050020
13050022 13050023 13050025 13050026 13050027 13050028 13050029 13050031 13050032 13050033
13050036 13050037 13050038 13050039 13050041 13050042 13050043 13050044 13050046 13050047
13050049 13050050 13051006 13051007 13051008 13051009 13051010 13051011 13051012 13051015
13051016 13051017 13051019 13051020 13051021 13051022 13051023 13051024 13051026 13051028
13051068 13051069 13051070 13051071 13051072 13051073 13051074 13051080 13051081 13051083
13051085 13051086 13051087 13051088 13051092 13051093 13051095 13052001 13052002 13052003
13052004 13052005 13052009 13052010 13052011 13052012 13052013 13052014 13052015 13052016

13052017 13052018 13052019 13052020 13052021 13052022 13052036 13052037 13052039 13052042
13052043 13052045 13052048 13052050 13052051 13052052 13052053 13052054 13052056 13052061
13052088 13053004 13053005 13053006 13053007 13053012 13053013 13053015 13053027 13053028
13054022 13054023 13054044 13054045 13054060 13054061 13054062 13054063 13054064 13054065
13054066 13054068 13054069 13054070 13054084 13054085 13055001 13055004 13055006 13055007
13055008 13055009 13055010 13055011 13055014 13055015 13055016 13055017 13055018 13055019
13055020 13055021 13055022 13055023 13055024 13055035 13055036 13055037 13055038 13055039
13055068 13055070 13055072 13055073 13055075 13055076 13055080 13055081 13055082 13055086
13055087 13056005 13056007 13056008 13056020 13056021 13056022 13056023 13056024 13056025
13056026 13056027 13056028 13056029 13056030 13056031 13056033 13056034 13056035 13056037
13056038 13056039 13057005 13057006 13057007 13057008 13057009 13057010 13057011 13057014
13057015 13057016 13057017 13057018 13057019 13057021 13057022 13057023 13057024 13057025
13057026 13057027 13057044 13057045 13057046 13057047 13057048 13057049 13057050 13057051
13057052 13057053 13057055 13057056 13057057 13057058 13058002 13058008 13058015 13058016
13058017 13058018 13058025 13058026 13058028 13058029 13058030 13058031 13058032 13059005
13059006 13059007 13059008 13059009 13059010 13059011 13059012 13059013 13059014 13059015
13059016 13059017 13059018 13059019 13059020 13059021 13059022 13059023 13059025 13059026
13059027 13059076 13059077 13059078 13059079 13059080 13059082 13059083 13059084 13059085
13059086 13059087 13060001 13060002 13060003 13060008 13060009 13060010 13060011 13060012
13061024 13061025 13061026 13061029 13061030 13061031 13061035 13061054 13061055 13061059
13061060 13061061 13062001 13062002 13062004 13062005 13062006 13062007 13062013 13062025
13062026 13062028 13062029 13062049 13062050 13062052 13062059 13062060 13062061 13062062
13062063 13063009 13063010 13063011 13063020 13063022 13063023 13063030 13063031 13063032
13063033 13063034 13063035 13063036 13063053 13063054 13063062 13063063 13063065 13063067
13063068 13063071 13063072 13063073 13063074 13063076 13064001 13064002 13064003 13064004
13064005 13064006 13064012 13064014 13064020 13064021 13064022 13064023 13064024 13064025
13064026 13064028 13064029 13064030 13064031 13064032 13064052 13064055 13064056 13064057
13064059 13064061 13064064 13064065 13064066 13064068 13064070 13064074 13064075 13065005
13065006 13065007 13065008 13065009 13065013 13065014 13065015 13065016 13065017 13065018
13065019 13065020 13065021 13065022 13065046 13065047 13065048 13065049 13065050 13065052
13065053 13065055 13065056 13065058 13065059 13065060 13066021 13066022 13066023 13066024
13066025 13066026 13066027 13066028 13066029 13066030 13066031 13066033 13066034 13066035
13066036 13066101 13066102 13066104 13066109 13066110 13067001 13067002 13067003 13067010
13067011 13067012 13067013 13067014 13067015 13067017 13068017 13068022 13068027 13068029
13068034 13068036 13068037 13068084 13068085 13068086 13068087 13068088 13068090 13069001
13069002 13069003 13069004 13069005 13069006 13069007 13069008 13069013 13069014 13069016
13069017 13069018 13069020 13069021 13069022 13069023 13069024 13069026 13069027 13069029
13069030 13069031 13069035 13069036 13069066 13069067 13069068 13069069 13069073 13070006
13070008 13070010 13070011 13070012 13070014 13070015 13070016 13070017 13070018 13070019
13070020 13070021 13070022 13070024 13070025 13070026 13070027 13070056 13070057 13070058
13070059 13070060 13070061 13071003 13071004 13071005 13071006 13071008 13071009 13071010
13071011 13071012 13071064 13072001 13072002 13072003 13072004 13072005 13072006 13072007
13072008 13072009 13072010 13072011 13072014 13072015 13072016 13072017 13072018 13072019
13072020

D   Run 2012 List (Embedding)

13047003 13047004 13047022 13047023 13047024 13047026 13047027 13047028 13047029 13047030
13047031 13047032 13047033 13047034 13047035 13047036 13047037 13047039 13047041 13047042
13047043 13047044 13047122 13047123 13047124 13047126 13048009 13048010 13048011 13048012
13048013 13048014 13048015 13048016 13048017 13048018 13048019 13048030 13048031 13048032
13048040 13048041 13048042 13048043 13048044 13048045 13048046 13048049 13048050 13048051
13048052 13048053 13048087 13048088 13048089 13048090 13048091 13048092 13048093 13049004
13049005 13049006 13049007 13049031 13049032 13049035 13049039 13049041 13049042 13049044
13049045 13049046 13049047 13049048 13049049 13049050 13049072 13049073 13049080 13049081
13049082 13049086 13049087 13049088 13049089 13049093 13049094 13049096 13049098 13049099
13049101 13050001 13050006 13050007 13050009 13050011 13050012 13050015 13050016 13050020
13050022 13050023 13050025 13050026 13050027 13050028 13050029 13050031 13050032 13050033
13050036 13050037 13050038 13050039 13050041 13050042 13050043 13050044 13050046 13050047
13050049 13050050 13051006 13051007 13051008 13051009 13051010 13051011 13051012 13051015
13051016 13051017 13051019 13051020 13051021 13051022 13051023 13051024 13051026 13051028
13051068 13051069 13051070 13051071 13051072 13051073 13051074 13051080 13051081 13051083
13051085 13051086 13051087 13051088 13051092 13051093 13051095 13052001 13052002 13052003
13052004 13052005 13052009 13052010 13052011 13052012 13052013 13052014 13052015 13052016
13052017 13052018 13052019 13052020 13052021 13052022 13052036 13052037 13052039 13052042
13052043 13052045 13052048 13052050 13052051 13052052 13052053 13052054 13052056 13052061
13052088 13053004 13053005 13053006 13053007 13053012 13053013 13053015 13053027 13053028
13054022 13054023 13054044 13054045 13054060 13054061 13054062 13054063 13054064 13054065
13054066 13054068 13054069 13054070 13054084 13054085 13055001 13055004 13055006 13055007
13055008 13055009 13055010 13055011 13055014 13055015 13055016 13055017 13055018 13055019
13055020 13055021 13055022 13055023 13055024 13055035 13055036 13055037 13055038 13055039
13055068 13055070 13055072 13055073 13055075 13055076 13055080 13055081 13055082 13055086
13055087 13056005 13056007 13056008 13056020 13056021 13056022 13056023 13056024 13056025
13056026 13056027 13056028 13056029 13056030 13056031 13056033 13056034 13056035 13056037
13056038 13056039 13057005 13057006 13057007 13057008 13057009 13057010 13057011 13057014
13057015 13057016 13057017 13057018 13057019 13057021 13057022 13057023 13057024 13057025
13057026 13057027 13057044 13057045 13057046 13057047 13057048 13057049 13057050 13057051
13057052 13057053 13057055 13057056 13057057 13057058 13058008 13058016 13058017 13058018
13058025 13058026 13058028 13058029 13058030 13058031 13058032 13059005 13059006 13059007
13059008 13059009 13059010 13059011 13059012 13059013 13059014 13059015 13059016 13059017
13059018 13059019 13059020 13059021 13059022 13059023 13059025 13059026 13059027 13068017
13068022 13068027 13068029 13068034 13068036 13068037 13068084 13068085 13068086 13068087
13068088 13068090 13069001 13069002 13069003 13069004 13069005 13069006 13069007 13069008
13069013 13069014 13069016 13069017 13069018 13069020 13069021 13069022 13069023 13069024
13069026 13069027 13069029 13069030 13069031 13069035 13069066 13069067 13069068 13069069
13069073 13070006 13070008 13070010 13070011 13070012 13070014 13070015 13070016 13070017
13070018 13070019 13070020 13070021 13070022 13070024 13070025 13070027 13070056 13070057
13070058 13070059 13070060 13070061 13071003 13071004 13071005 13071006 13071008 13071009
13071010 13071011 13071012 13071064 13072001 13072003 13072004 13072005 13072006 13072007
13072008 13072009 13072010 13072011 13072014 13072015 13072016 13072017 13072018 13072019
13072020