



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Examination of flow and nonflow factorization methods in small collision systems

S. H. Lim, Q. Hu, R. Belmont, K. K. Hill, J. L. Nagle, and D. V. Perepelitsa

Phys. Rev. C **100**, 024908 — Published 16 August 2019

DOI: [10.1103/PhysRevC.100.024908](https://doi.org/10.1103/PhysRevC.100.024908)

Examination of Flow and Non-Flow Factorization Methods in Small Collision Systems

S.H. Lim,¹ Q. Hu,¹ R. Belmont,² K.K. Hill,¹ J.L. Nagle,¹ and D.V. Perepelitsa¹

¹*University of Colorado, Boulder, Colorado 80309, USA*

²*University of North Carolina, Greensboro, North Carolina 27413, USA*

(Dated: July 15, 2019)

Two particle correlations have been used extensively to study hydrodynamic flow patterns in heavy-ion collisions. In small collision systems, such as $p+p$ and $p+A$, where particle multiplicities are much smaller than in $A+A$ collisions, non-flow effects from jet correlations, momentum conservation, particle decays, etc. can be significant, even when imposing a large pseudorapidity gap between the particles. A number of techniques to subtract the non-flow contribution in two particle correlations have been developed by experiments at the Large Hadron Collider (LHC) and then used to measure particle flow in $p+p$ and $p+Pb$ collisions. Recently, experiments at the Relativistic Heavy Ion Collider (RHIC) have explored the possibility of adopting these techniques for small collision systems at lower energies. In this paper, we test these techniques using Monte Carlo generators PYTHIA and HIJING, which do not include any collective flow, and AMPT, which does. We find that it is crucial to examine the results of such tests both for correlations integrated over particle transverse momentum p_T and differentially as a function of p_T . Our results indicate reasonable non-flow subtraction for $p+p$ collisions at the highest LHC energies, while failing if applied to $p+p$ collisions at RHIC. In the case of $p+Au$ collisions at RHIC, both HIJING and AMPT results indicate a substantial over-subtraction of non-flow for $p_T \gtrsim 1$ GeV/ c and hence an underestimate of elliptic flow.

I. INTRODUCTION

The standard evolution model of relativistic heavy ion collisions involves the translation of initial geometric anisotropies into final momentum correlations via nearly perfect hydrodynamic flow followed by hadronic re-scattering [1, 2]. Starting with first indications in $p+p$ collisions at the LHC [3], there is now a wealth of data indicating similar translation of geometry into flow in smaller collision systems at RHIC and the LHC [4]. Here we focus on $p+p$ collisions at the LHC and $p+p$ and $p+Au$ collisions at RHIC where one is pushing the limits of how small a system and how small a particle multiplicity still results in significant final flow signatures. It is imperative to consider both collision energies and systems since a simultaneous description within the perfect fluid paradigm currently gives the best data description [5, 6].

In these small systems, there are many contributions to the final particle correlations and some of these contributions increase in relative strength with decreasing multiplicity. There are many sources of correlations among only a subset of the final particles that result in an increased probability of particles to be nearby in momentum space (i.e. close in both pseudorapidity $\Delta\eta \approx 0$ and azimuthal angle $\Delta\phi \approx 0$). Examples include the decay of heavy resonances, correlations due to quantum statistics (HBT effects) [7], and the fragmentation of a high momentum quark or gluon—the last of these resulting in what are referred to as jet correlations. These correlations are effectively minimized by requiring the two particles to have a substantial gap in pseudorapidity, typically $|\Delta\eta| > 2$. There are other correlations that survive such a cut including di-jet correlations. In a leading-order hard scattering, a parton (quark or gluon) and partner parton will be nearly back-to-back in azimuth ($\Delta\phi \approx \pi$)

and can be widely separated in pseudorapidity, since the incoming partons need not be longitudinally momentum balanced. As such, one may expect two particle correlations with an enhancement near $\Delta\phi \approx \pi$ that extends long-range in pseudorapidity.

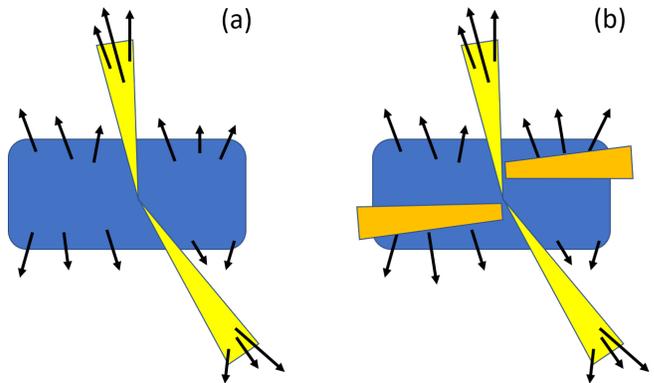


FIG. 1. Panel (a): A cartoon of the simplest scenario with medium particles emitted from the bulk (blue region) and jet particles emitted from two fragmenting partons (yellow cones). Panel (b): A cartoon with the addition of two longitudinally oriented strings such that momentum conservation effects may be present over large rapidity regions.

As shown in Figure 1, one can imagine a toy scenario where a medium indicated by the blue region is created in a collision. In addition, a hard scattering results in two partons that largely fragment into hadrons outside of the medium. In a perfectly factorized picture, the medium particles may have angular correlations that reflect the flowing fluid while the jet particles are correlated with each other but are uncorrelated with the medium geometry. As one examines collisions with larger final particle

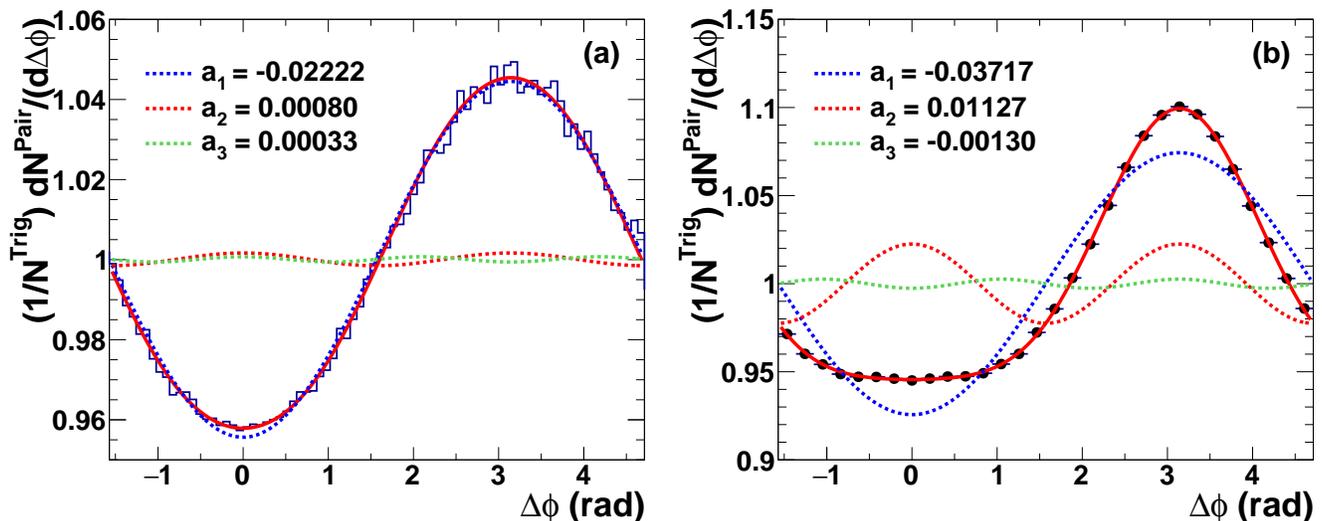


FIG. 2. Panel (a): Two particle correlation for pairs generated from pure N -body phase space with a pseudorapidity gap $|\Delta\eta| > 2$ requirement. Panel (b): Two particle correlation for pairs generated in PYTHIA8 with a pseudorapidity gap $|\Delta\eta| > 2$ requirement. Fit values for the Fourier decomposition (a_1 , a_2 , and a_3) are also shown.

multiplicities and larger volume, a larger fraction of the particles come from the medium and thus the influence of the jet correlations is reduced. In this simplest scenario the correlations from the jet particles are referred to as “non-flow” and those from the medium particles are referred to as “flow.” We highlight that it is impossible for this factorization to be perfectly true. The jet partons cannot hadronize without some color interaction with the medium or underlying event, and in some case many of the jet partons scatter with medium partons. For that reason there is no direct separability of the two, rather they are related through a convolution.

Another key source of non-flow correlations is simply global momentum conservation. The simplest case would be to imagine N particles emitted isotropically but obeying momentum conservation. The ROOT software package [8] has a class TGenPhaseSpace that allows for N -body decays simply following phase space filling and momentum conservation.

Of course, we do not expect that the N particles are distributed simply according to phase space rules. Within a model such as PYTHIA8 [9], one can think of every two-jet event as really being at least a four-jet event. As shown in panel b of Figure 1, if two incoming partons have a large momentum transfer, the resulting beam remnants extend nearly longitudinal color strings (i.e. the other two jets). These strings can have a transverse momentum kick such that particles emitted from the upper (orange) string have a slight trend upward and from the lower (orange) string a slight trend downward. Thus, the partons and resulting hadrons will have a momentum conservation correlation that may be convolved with a flow correlation if these particles undergo additional final state scattering.

The $\Delta\phi$ correlation function with the requirement

$|\Delta\eta| > 2$ is shown in Figure 2, where the N -body phase space calculation is shown in the left panel and the jet-type correlation from PYTHIA8 is shown in the right panel. Most correlation analyses characterize the two particle correlation in $\Delta\phi$ in terms of a Fourier decomposition [10]:

$$f(\Delta\phi) = G \left\{ 1 + 2 \sum_{n=1}^{\infty} a_n \cos(n\Delta\phi) \right\}, \quad (1)$$

where the coefficients a_n are the Fourier coefficient at order n , G is the normalization factor corresponding to the average number of associated particles per trigger particle in the sample, and all sine terms vanish due to symmetry. The coefficients a_n are shown for the two non-flow contributions in Figure 2. In the case of the N -body phase space, the dominant contribution is c_1 with a depletion near $\Delta\phi \approx 0$ and an enhancement of pairs near $\Delta\phi \approx \pi$. However, there are non-zero higher order contributions and their relative strengths will of course depend on the other correlations embedded along with momentum conservation. For the jet-type correlations, the dominant term is also a_1 due to the strong long range away-side peak. However, the successive terms contribute significantly and with alternating signs in order to describe the nearly flat region around $\Delta\phi \approx 0$ and the peak around $\Delta\phi \approx \pi$.

In the following sections we define the non-flow subtraction methods and then detail the resulting tests of these methods using the Monte Carlo models PYTHIA [9], HIJING [11], and AMPT [12] in different collisions systems and at different energies.

II. DEFINITION OF METHODS

The most important assumption in all non-flow subtraction methods is the assumption that the correlation coefficients a_n in Eq. 1 can be separated into two linearly additive contributions $a_n = c_n + d_n$, where c_n is the flow coefficient quantifying correlations related to the initial geometry and d_n is the non-flow coefficient of pair correlations. As discussed in Section I, this assumption may not be realistic, but this is the starting point for all non-flow subtraction methods, and so we begin our discussion here. With this assumption, Eq. 1 can be rewritten as:

$$f(\Delta\phi) = G \left\{ 1 + 2 \sum_{n=1}^{\infty} (c_n + d_n) \cos(n\Delta\phi) \right\}. \quad (2)$$

In order to extract the true flow, c_n , it is necessary to remove the non-flow contributions, d_n . This is particularly important in small systems where the non-flow correlations can dominate the flow signal. We can define the non-flow contribution $J(\Delta\phi)$ to $f(\Delta\phi)$ as:

$$J(\Delta\phi) = G \left\{ 2 \sum_{n=1}^{\infty} d_n \cos(n\Delta\phi) \right\}, \quad (3)$$

so that Eq. 2 becomes

$$f(\Delta\phi) = J(\Delta\phi) + G \left\{ 1 + 2 \sum_{n=1}^{\infty} c_n \cos(n\Delta\phi) \right\}. \quad (4)$$

All techniques attempting to disentangle flow and non-flow operate by comparing correlations between two data selection samples: one at higher multiplicity (HM), where flow is expected to have a larger influence; and one at lower multiplicity (LM), where flow is expected to have a smaller influence. The flow in LM events is at times assumed to be negligible or non-existent; however, we need not consider such differences just yet. The above equations simply need to be trivially relabeled so that the LM and HM categories are distinct. Here we use LM (HM) as a superscript on all relevant quantities to indicate LM (HM) events.

The terms c_n^{HM} and c_n^{LM} are the flow correlation coefficients at low and high multiplicity, respectively; d_n^{HM} and d_n^{LM} are the non-flow correlation coefficients at low and high multiplicity, respectively.

Applying these labels and rewriting Eq. 3 we obtain

$$\frac{J^{\text{LM(HM)}}(\Delta\phi)}{G^{\text{LM(HM)}}} = 2 \sum_{n=1}^{\infty} d_n^{\text{LM(HM)}} \cos(n\Delta\phi), \quad (5)$$

meaning the $J^{\text{LM}}(\Delta\phi)$ and $J^{\text{HM}}(\Delta\phi)$ are related to each other as:

$$\frac{J^{\text{HM}}(\Delta\phi)}{G^{\text{HM}}} = \frac{\sum_{n=1}^{\infty} j_n d_n^{\text{LM}} \cos(n\Delta\phi)}{\sum_{n=1}^{\infty} d_n^{\text{LM}} \cos(n\Delta\phi)} \frac{J^{\text{LM}}(\Delta\phi)}{G^{\text{LM}}}, \quad (6)$$

where the relational coefficients are $j_n = d_n^{\text{HM}}/d_n^{\text{LM}}$. By estimating the j_n and the flow coefficients at low multiplicity, c_n^{LM} , the non-flow contribution at high multiplicity can be determined and subtracted. There are a few different non-flow subtraction methods available on the market based on different assumptions regarding the c_n^{LM} and j_n coefficients.

A. Method 1

The first method we detail has been developed by the ATLAS Collaboration, as originally applied in Ref. [13]. Here we detail the key assumptions made in using this method. The first assumption is that the non-flow correlation coefficient $j_n = d_n^{\text{HM}}/d_n^{\text{LM}}$ is independent of the harmonic number n , so that $j_n = j^{\text{ATLAS}}$ for all n . This allows Eq. 6 to be rewritten as:

$$\frac{J^{\text{HM}}(\Delta\phi)}{G^{\text{HM}}} = j^{\text{ATLAS}} \frac{J^{\text{LM}}(\Delta\phi)}{G^{\text{LM}}}. \quad (7)$$

This assumption follows from the idea that at the LHC the non-flow correlation is dominated by the jet contribution. The additional requirement is then that the shape of the jet correlation does not change with multiplicity event class.

A second assumption made is that the flow contribution at the lowest harmonic c_1 is negligible compared to the non-flow d_1 , i.e. $c_1 \ll d_1$. Hence c_1 is completely ignored; alternately one can think of it as being absorbed via redefinition of d_1 . Following these assumptions, we can rewrite the angular correlation distributions in Eq. 4 for HM events in relation to LM events:

$$f^{\text{LM}}(\Delta\phi) = J^{\text{LM}}(\Delta\phi) + G^{\text{LM}} \left\{ 1 + 2 \sum_{n=2}^{\infty} c_n^{\text{LM}} \cos(n\Delta\phi) \right\}, \quad (8)$$

$$f^{\text{HM}}(\Delta\phi) = J^{\text{HM}}(\Delta\phi) + G^{\text{HM}} \left\{ 1 + 2 \sum_{n=2}^{\infty} c_n^{\text{HM}} \cos(n\Delta\phi) \right\}, \quad (9)$$

and then combining the two using Eq. 7 to obtain

$$f^{\text{HM}}(\Delta\phi) = \frac{G^{\text{HM}} j^{\text{ATLAS}}}{G^{\text{LM}}} f^{\text{LM}}(\Delta\phi) + G^{\text{HM}} (1 - j^{\text{ATLAS}}) \left\{ 1 + 2 \sum_{n=2}^{\infty} \left(\frac{c_n^{\text{HM}} - j^{\text{ATLAS}} c_n^{\text{LM}}}{1 - j^{\text{ATLAS}}} \right) \cos(n\Delta\phi) \right\} \quad (10)$$

$$= F^{\text{temp}} f^{\text{LM}}(\Delta\phi) + G^{\text{temp}} \left\{ 1 + 2 \sum_{n=2}^{\infty} c_n^{\text{temp}} \cos(n\Delta\phi) \right\}, \quad (11)$$

where F^{temp} , G^{temp} , and c_n^{temp} are parameters defined by

$$F^{\text{temp}} = \frac{G^{\text{HM}} j^{\text{ATLAS}}}{G^{\text{LM}}}, \quad (12)$$

$$G^{\text{temp}} = G^{\text{HM}} (1 - j^{\text{ATLAS}}), \quad (13)$$

$$c_n^{\text{temp}} = \frac{c_n^{\text{HM}} - j^{\text{ATLAS}} c_n^{\text{LM}}}{1 - j^{\text{ATLAS}}}. \quad (14)$$

These parameters are obtained from fitting $f^{\text{HM}}(\Delta\phi)$ with $f^{\text{LM}}(\Delta\phi)$. Thus, this approach is often referred to as a ‘‘template-fitting’’ method because the LM correlation function serves as a template for the HM correlation function.

In the special case where the flow coefficients are identical at all orders between the low multiplicity and high multiplicity classes, i.e. $c_n^{\text{HM}} = c_n^{\text{LM}}$ for all n , then the extracted c_n^{temp} is exactly equal to c_n^{HM} . This is a simplifying assumption that has no concise physics motivation. This special case was assumed in the ATLAS publications [13, 14].

Relaxing this special case, an additional correction should be applied to the fitted value to obtain c_n^{HM} :

$$c_n^{\text{HM}} = c_n^{\text{temp}} - j^{\text{ATLAS}} (c_n^{\text{temp}} - c_n^{\text{LM}}), \quad (15)$$

where j_n can be obtained using fit parameters $j^{\text{ATLAS}} = G^{\text{LM}} F^{\text{temp}} / G^{\text{HM}}$ and c_n^{LM} is usually estimated using the flow coefficient measured in the second lowest multiplicity sample. The ATLAS Collaboration applies this additional correction in Ref. [15].

B. Method 2

The second method we detail has been developed by the CMS Collaboration, as originally applied in Ref. [16]. Similar approaches have also been explored by the ATLAS and ALICE collaborations [17, 18]. Here we detail the key assumptions made in using this method.

Like the ATLAS method, this method also assumes that the non-flow correlation shape does not change with multiplicity class, hence $j_n = a_n^{\text{HM}} / a_n^{\text{LM}}$ is independent of the harmonic number n , so that $j_n = j^{\text{CMS}}$ for all n . This allows Eq. 6 to be rewritten as:

$$\frac{J^{\text{HM}}(\Delta\phi)}{G^{\text{HM}}} = j^{\text{CMS}} \frac{J^{\text{LM}}(\Delta\phi)}{G^{\text{LM}}}, \quad (16)$$

in exact analog to the ATLAS case. However, the method by which j^{CMS} is determined is different. They assume that the non-flow correlation coefficients, at least for $n > 1$, are dominated by jet contributions. Since they assume the jet shape is independent of multiplicity class, they only need to determine the relative jet yields in the different multiplicity classes. They do this by measuring the near-side jet yield in the different multiplicity classes via the short range (SR) correlation, i.e. with $\Delta\phi \approx 0$ and $\Delta\eta \approx 0$. They isolate the near-side jet peak by taking the difference between the SR and long range (LR) correlations as follows:

$$j^{\text{CMS}} = \frac{G^{\text{LM}} \int_{-1.2}^{1.2} (f_{\text{SR}}^{\text{HM}}(\Delta\phi) - f_{\text{SR}}^{\text{LM}}(\Delta\phi)) d\Delta\phi}{G^{\text{HM}} \int_{-1.2}^{1.2} (f_{\text{SR}}^{\text{LM}}(\Delta\phi) - f_{\text{LR}}^{\text{LM}}(\Delta\phi)) d\Delta\phi}, \quad (17)$$

where $f_{\text{SR}}^{\text{HM(LM)}}(\Delta\phi)$ is the short-range correlation with $|\Delta\eta| < 1$ and $f_{\text{LR}}^{\text{HM(LM)}}(\Delta\phi)$ is the long-range correlation with $2 < |\Delta\eta| < 5$. The short-range and long-range correlations are usually normalized by number of trigger particles such that the integral of their difference over $-1.2 < \Delta\phi < 1.2$ is close to the number of particles produced per jet.

Next, direct Fourier fits are applied to the long-range correlations at low and high multiplicities, and the Fourier coefficient $a_n^{\text{HM(LM)}}$ are extracted. This approach assumes that there is no flow contribution at all in the low multiplicity event selection, i.e. $c_n^{\text{LM}} = 0$. This is an extreme assumption that forces any extraction of flow coefficients as a function of multiplicity class to approach zero at low multiplicity, and is quite different from the ATLAS approach. Then at each order n the flow coefficient at high multiplicity is given by:

$$c_n^{\text{HM}} = a_n^{\text{HM}} - j^{\text{CMS}} a_n^{\text{LM}}. \quad (18)$$

As noted above, as measurements move towards low multiplicity, such that $c_n^{\text{HM}} \approx c_n^{\text{LM}}$, c_n^{HM} will converge to zero by construction. The CMS results using this method thus always trend to zero. However, if a non-zero flow coefficient in the low multiplicity sample is allowed, Eq. 18 could be rewritten as:

$$c_n^{\text{HM}} = a_n^{\text{HM}} - j^{\text{CMS}} (a_n^{\text{LM}} - c_n^{\text{LM}}), \quad (19)$$

where c_n^{LM} can be estimated in the same way as in the ATLAS method. Once the generalized results as shown in Eq. 15 and Eq. 18 are used, one should be able to obtain similar results with the two methods.

One remaining difference between the methods would arise from the different ways of estimating the jet variable j_n . The non-flow correlation can have contributions from jets as well as overall momentum conservation. The momentum conservation contribution is predominantly in the d_1 coefficient. Thus, in the CMS method where they subtract the non-flow order by order, even though the j^{CMS} is determined from jets alone, it does not matter since they are not extracting a first order c_1 at high multiplicity. In contrast, in the ATLAS template-fitting method, it assumes the non-flow shape (combining both jets and momentum conservation) scale in the same way. This seems unlikely given the findings by CMS via the short range correlations. Since neither experiment extracts a c_1 at high multiplicity where they would get very different results, there is only the potential for a residual effect on the higher order c_n via the ATLAS fitting procedure.

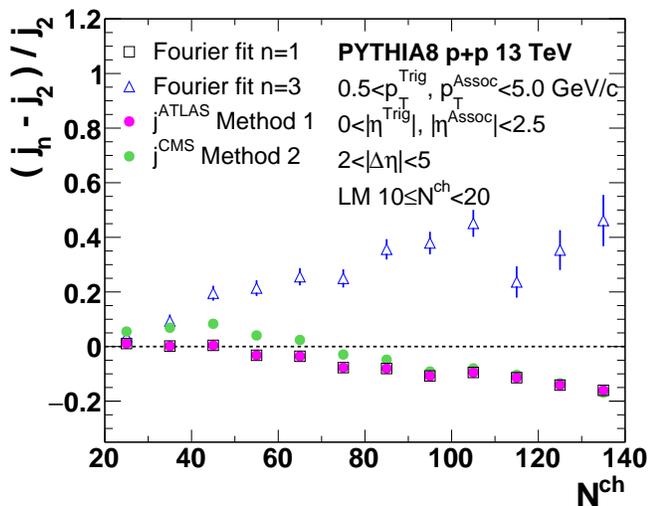


FIG. 3. Relative difference between the ratio of non-flow correlation coefficient at high multiplicity to that at low multiplicity, $j_n = d_n^{\text{HM}}/d_n^{\text{LM}}$, from different methods to that from direct Fourier fit of the second order, j_2 , as a function of multiplicity. A pseudorapidity gap of $2 < |\Delta\eta| < 5$ is required.

C. Jet Shape Assumption

Method 1 and *2* are derived from the same expression and heavily rely on the assumption that the jet shape is the same at low and high multiplicity. This assumption can be tested in PYTHIA8, in which no flow contribution is expected, i.e. $c_n = 0$. We perform a direct Fourier coefficient extraction in the low and high multiplicity event classes and then compute $j_n = d_n^{\text{HM}}/d_n^{\text{LM}}$. To examine if all j_n are the same, we plot $(j_n - j_2)/j_2$ as shown in Figure 3. The points for $n = 1$ and $n = 3$ indicate that the assumption of a common j_n is violated in PYTHIA8 and that violation increases with higher multiplicity, reach-

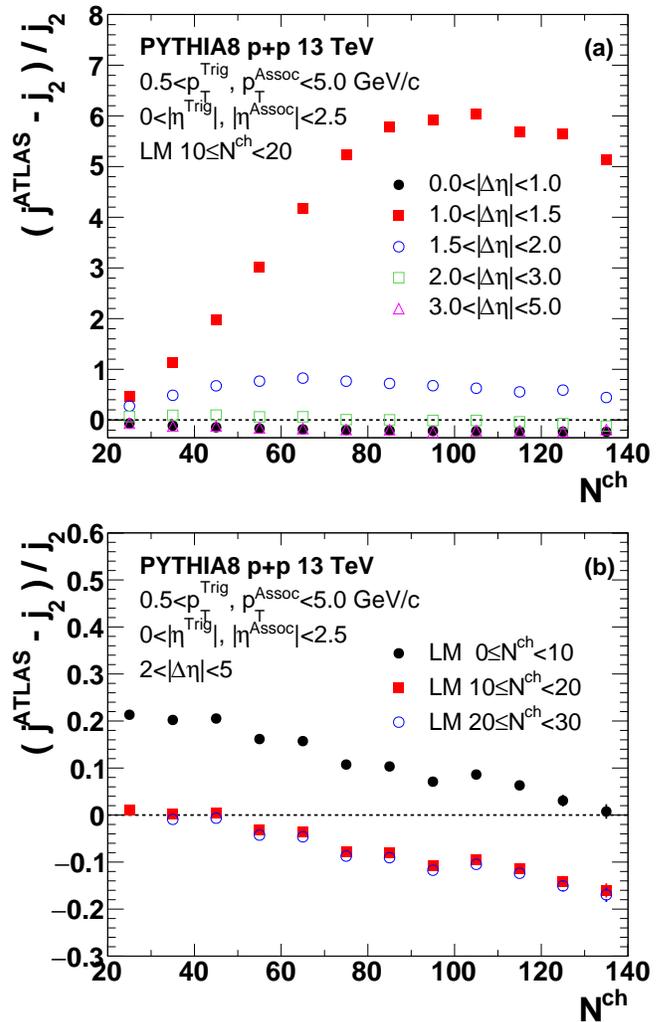


FIG. 4. The relative difference between j^{ATLAS} and j_2 from direct Fourier fit (a) as a function of multiplicity for different pseudorapidity gaps and (b) different choices of low multiplicity sample.

ing a level for $n = 1$ ($n = 3$) relative to $n = 2$ of -20% ($+40\%$). We note that the j_2 is decreasing with multiplicity, so the impact of this violation might be actually smaller at higher multiplicity. We also show the j^{ATLAS} and j^{CMS} extracted values. One can see that they closely agree with the $n = 1$ case and deviate significantly from the $n = 3$ case. Note that the j^{ATLAS} is actually identical to the $n = 1$. This is because there is no first order flow coefficient ($c_1 = 0$) by construction. The fit procedure results in j^{ATLAS} being determined completely by the first order non-flow correlation coefficient (d_1), while the contribution from higher order non-flow correlation coefficients is absorbed into the higher-order flow correlation coefficients.

We note that the j_n assumption can also be sensitive to the particular pseudorapidity gap chosen and the low multiplicity reference selection. The sensitivity of the j_n estimated by *Method 1* to the choice of pseudorapidity

gap and low multiplicity selection is shown in Figure 4. The method has particular problems when the pseudorapidity gap is $1 < |\Delta\eta| < 1.5$, because there are contributions from both the near and away-side jet. The method also is most sensitive when the low multiplicity selection is at its lowest.

All of these violations of the assumptions in *Method 1* and *2* can only be gauged in terms of consequences on the extracted flow coefficients c_n by testing the methods on various Monte Carlo physics models. In the following sections, we examine the results of such tests.

III. CLOSURE TESTS WITH PYTHIA8 AND HIJING

We now test these procedures on Monte Carlo generators. In Monte Carlo generators such as PYTHIA8 [9] and HIJING [11], there is no collective flow in a sense that there are no final state interactions to translate a spatial geometry into momentum anisotropies. Thus, one expects that the application of a successful non-flow subtraction method should result in flow anisotropy coefficients of exactly zero in these cases. These are thus referred to as closure tests. By measuring the residual of these coefficients c_n , the level of closure can be quantified. We have applied the methods described in Section II to determine the level of closure in various Monte Carlo generators.

Before proceeding we want to define the nomenclature used in the following sections. In our studies we focus on the extraction of elliptic flow $n = 2$, and show all results in terms of v_{22} and v_{22}^{sub} extracted from the two particle correlations. First we define $v_{22} = a_2(p_{T,1}, p_{T,2})$, where the trigger and associated particles are in momentum selections around $p_{T,1}$ and $p_{T,2}$, respectively. Once the non-flow subtraction technique is applied and the c_n coefficients are estimated, we define $v_{22}^{\text{sub}} = c_2(p_{T,1}, p_{T,2})$. These quantities are related but not equal to v_2^2 . The standard differential v_2 as a function of p_T would then be extracted as $v_2(p_{T,1}) = v_{22}/\sqrt{a_2(p_{T,2}, p_{T,2})}$ and $v_2^{\text{sub}}(p_{T,1}) = v_{22}^{\text{sub}}/\sqrt{c_2(p_{T,2}, p_{T,2})}$. In the case where both $p_{T,1}$ and $p_{T,2}$ represent the same broad range in p_T , e.g. $0.5 < p_T < 5$ GeV/c, these expressions reduce to $v_2 = \sqrt{v_{22}} = \sqrt{a_2}$ and $v_2^{\text{sub}} = \sqrt{v_{22}^{\text{sub}}} = \sqrt{c_2}$, representing integral v_2 as a function of some event-level variable (multiplicity, centrality, etc.).

We also highlight that experiments have different techniques for selecting lower and higher multiplicity events. If the selection is based on charged particle multiplicity centered around midrapidity, i.e. in the same range as the two particles used for the correlations, we label the event categories in terms of N^{ch} . In contrast, other measurements utilize multiplicity or energy in a forward or backward rapidity range, for example in the Pb-going direction in p +Pb collisions, and thus outside the range of the two particles used for the correlations. In this second case, we refer to the event selections by ‘‘centrality.’’

A. LHC $p+p$ Case

These methods were developed particularly by the ATLAS and CMS Collaborations for use in the highest energy $p+p$ collisions, and we examine that collision system first. We note that the CMS Collaboration has presented the results of a PYTHIA8 closure test for particles integrated over p_T in their paper (see Figure 4 of Ref. [16]), but such a study has not been published by ATLAS. We consider similar acceptance selection cuts as the experiments, though use one set so that all methods are compared apples-to-apples. Correlations are determined from all charged hadrons within $|\eta| < 2.5$ from 100 million PYTHIA8 $p+p$ events (`SoftQCD:nonDiffractive = on`) at $\sqrt{s} = 13$ TeV. Charged hadrons with $p_T > 0.4$ GeV/c ($p_T > 0.5$ GeV/c) are used for event multiplicity categorization (particle correlations). We note that a slightly different p_T selection ($p_T > 0.3$ GeV/c) has been used for data analysis by the CMS collaboration.

Figure 18 (shown in Appendix A for clarity) shows one dimensional $\Delta\phi$ two particle correlation functions for short-range ($|\Delta\eta| < 1$) and long-range ($2 < |\Delta\eta| < 5$) regions. Charged hadrons are required to satisfy $p_T > 0.5$ GeV/c and $|\eta| < 2.5$. Each panel presents a different range of charged hadron multiplicity event selection N^{ch} . In this case, there is no visually obvious shape variation of the $\Delta\phi$ correlation functions both in short and long ranges throughout the entire multiplicity range. This implies that the j_n values are going to be approximately independent of n , though not exactly as demonstrated in the previous discussion. The dashed lines are fits to the distributions of long-range correlations to extract Fourier coefficients described in Eq. 1.

Figure 5 shows the second order Fourier coefficients v_{22} extracted directly from the correlation functions and after the non-flow subtraction technique is applied. Results are shown for charged hadrons with $0.5 < p_T < 5$ GeV/c as a function of event multiplicity N^{ch} . Figure 6 show the results for the highest multiplicity selection and for trigger particles as a function of p_T . The left and right panels of both figures apply the *Method 1* (ATLAS) and *Method 2* (CMS) methods, respectively.

The non-zero value of the coefficients after subtraction indicate that the closure test is not perfectly satisfied. This is mainly due to remaining jet correlations from the near-side ($\Delta\phi \approx 0$), even with the large $\Delta\eta$ gap, and/or a small shape change from the away-side ($\Delta\phi \approx \pi$). The non-flow effect on v_{22} is larger at low multiplicity events, and it becomes smaller as the event multiplicity increases. The two non-flow subtraction methods are applied with three different ranges of low multiplicity selection, and the results of the subtracted v_{22}^{sub} from each method are shown in each panel.

In Figure 5 the gray bands correspond to $|v_2| = 0.03$ (thus equivalent to $v_{22} = 0.03 \times 0.03$). There is some sensitivity to the low multiplicity selection. The v_{22}^{sub} with low multiplicity selection $10 \leq N^{\text{ch}} < 20$ and $20 \leq$

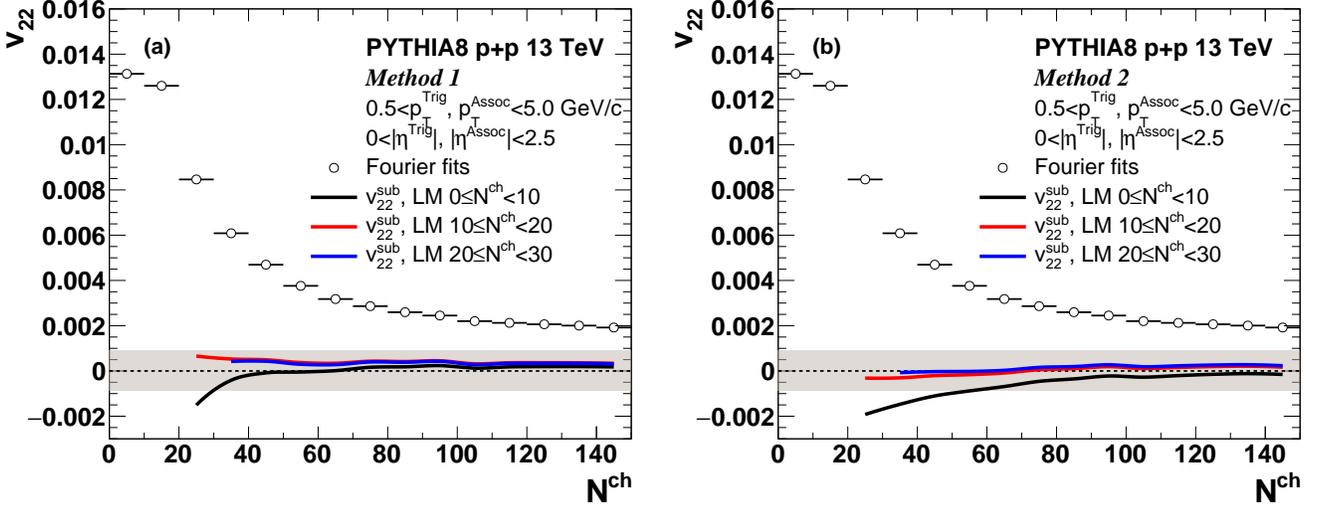


FIG. 5. The second order Fourier coefficient v_{22} of long-range ($2 < |\Delta\eta| < 5$) two particle correlation as a function of charged hadron multiplicity in $p+p$ collisions at $\sqrt{s} = 13$ TeV from PYTHIA8 before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.4$ GeV/c and $|\eta| < 2.5$. Gray bands correspond to a 3% $|v_2|$ window.

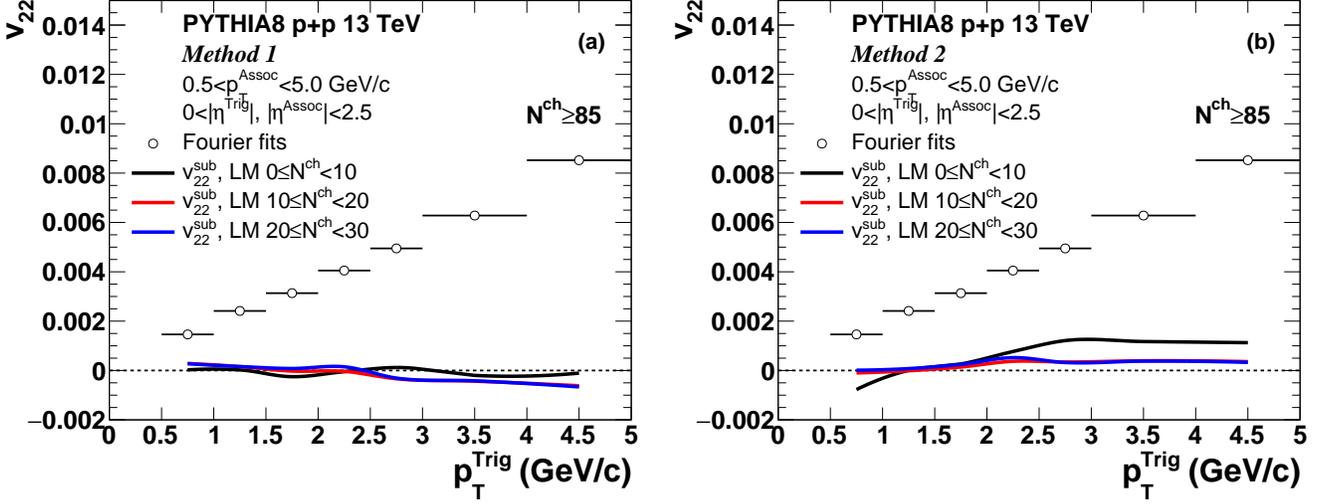


FIG. 6. The second order Fourier coefficient v_{22} of long-range ($2 < |\Delta\eta| < 5$) two particle correlation as a function of p_T in $p+p$ collisions of $N^{\text{ch}} \geq 85$ at $\sqrt{s} = 13$ TeV from PYTHIA8 before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.4$ GeV/c and $|\eta| < 2.5$.

$N^{\text{ch}} < 30$ are within the level of $|v_2| < 0.03$ over the entire multiplicity range. Using events in $0 \leq N^{\text{ch}} < 10$ as a low multiplicity bin results in v_{22}^{sub} with a slightly larger deviation from zero in the lower multiplicity range, but it converges with the v_{22}^{sub} from the other cases at higher multiplicities.

The same test is done as a function of the p_T of the trigger particle (p_T^{Trig}) for the highest 5% multiplicity events ($N^{\text{ch}} \geq 85$). Figure 19 in Appendix A shows the $\Delta\phi$ correlation functions for low ($0 \leq N^{\text{ch}} < 10$) and high ($N^{\text{ch}} \geq 85$) multiplicity events. Each panel shows a different p_T^{Trig} range, and the p_T range of associated particles is $0.5 < p_T^{\text{Assoc}} < 5$ GeV/c. The shape of the

near-side peak in the short-range correlation function becomes narrower as p_T^{Trig} increases, and there is no significant shape difference between low and high multiplicity events at the same p_T^{Trig} .

Figure 6 shows the v_{22} of long-range $\Delta\phi$ correlations as a function of p_T^{Trig} . Here, the v_{22} increases with p_T^{Trig} indicating a stronger non-flow effect at higher p_T^{Trig} . The lines represent v_{22}^{sub} using events from three different multiplicity ranges for the non-flow subtraction. As with the non-flow results as a function of multiplicity, the v_{22}^{sub} with reference events in $10 \leq N^{\text{ch}} < 20$ and $20 \leq N^{\text{ch}} < 30$ are quite consistent, and the v_{22}^{sub} values are smaller than

0.001 in $0.5 < p_T^{\text{Trig}} < 5$ GeV/c. The v_{22}^{sub} with events of the lowest multiplicity range ($0 \leq N^{\text{ch}} < 10$) is slightly different from the other two cases and shows a greater degree of non-closure at higher p_T using *Method 2*.

To summarize the closure test in $p+p$ collisions at $\sqrt{s} = 13$ TeV from PYTHIA8:

- The $\Delta\phi$ two particle correlation functions in PYTHIA8 exhibit minor violations of the assumptions in the non-flow subtraction methods examined.
- The PYTHIA8 resulting v_{22}^{sub} using *Method 1* and *2* pass the closure test much better than $|v_2| < 0.03$ as long as one avoids the lowest multiplicity range for the reference selection.
- The modest degree of non-closure with PYTHIA8 may be considered as a systematic uncertainty on the final extracted v_2 results.

B. RHIC $p+p$ Case

There are currently no measurements of extracted flow coefficients in $p+p$ collisions at RHIC. These are challenging measurements due to multiple collision pileup, much lower multiplicities compared with LHC collisions, and the more limited phase space acceptance of the RHIC experiments. However, we include this case for completeness and to inform future studies. For this test, we use all charged hadrons within $|\eta| < 2.5$ from one billion of PYTHIA8 $p+p$ events (`SoftQCD:nonDiffractive = on`) at $\sqrt{s} = 200$ GeV, and charged hadrons in $p_T > 0.2$ GeV/c are used for event multiplicity categorization. For these studies, we have modeled a very large acceptance similar to that of the LHC experiments.

Two particle $\Delta\phi$ and $\Delta\eta$ correlation functions are made with the same definition introduced in Section III A. One-dimensional $\Delta\phi$ correlation functions in short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) ranges in various multiplicity selections are presented in Figure 20 in Appendix A. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 2.5$ are used for the correlation function.

One obvious difference from the $\Delta\phi$ correlation functions of the LHC case (Figure 18) is that the shape of short-range $\Delta\phi$ correlation function significantly changes in the multiplicity range of $0 \leq N^{\text{ch}} < 40$ at $\sqrt{s} = 200$ GeV. In the case of the lowest multiplicity bin, ($0 \leq N^{\text{ch}} < 10$), the per-trigger-yield in the short-range correlation is smaller than that in the long-range correlation, so *Method 2* is simply not applicable within this multiplicity bin because it results in a negative jet yield.

Figure 7 shows the v_{22} from Fourier fits to the long-range $\Delta\phi$ correlation functions plotted versus multiplicity N^{ch} . Additionally plotted as solid lines are the v_{22}^{sub} values using three different low multiplicity reference ranges. The gray bands correspond to a window

of $|v_2| = 0.03$. In results from *Method 1* shown in the left panel. The v_{22}^{sub} clearly depends on the selection of low multiplicity reference, and the v_{22}^{sub} greatly deviate from zero in lower multiplicity ranges. This is because of the dramatic shape variation of the $\Delta\phi$ correlations in events of $0 \leq N^{\text{ch}} < 40$. When using events in the $20 \leq N^{\text{ch}} < 30$ multiplicity range as reference, the v_{22}^{sub} is within the window of $|v_2| = 0.03$.

In the right panel of Figure 7, the non-flow subtraction results using *Method 2* are presented; however, the lowest multiplicity bin ($0 \leq N^{\text{ch}} < 10$) is not included due to the negative jet yield as discussed earlier. The v_{22}^{sub} from *Method 2* using the remaining reference multiplicity ranges are within the window of $|v_2| = 0.03$. One possible reason for these results differing from *Method 1* is an interplay of the shape variation of $\Delta\phi$ correlation functions at both short and long range. In *Method 1*, the particular multiplicity-dependence of the long-range correlation function results in a large deviation of the v_{22}^{sub} from zero. However in *Method 2*, which also uses information about the scaling of the jet yield at short range, this effect is partially compensated and the deviation is significantly smaller. As shown in Figure 20, the difference of minimum per-trigger yields in short-range and long-range correlations strongly changes with multiplicity, and the additional jet yield in low multiplicity events possibly results in a smaller scaling for v_{22} of the low multiplicity events.

The outcome of the non-flow subtraction as a function of p_T in $p+p$ collisions at $\sqrt{s} = 200$ GeV has been studied as well. $\Delta\phi$ correlation functions with different p_T^{Trig} ranges at short and long ranges are presented in Figure 21 in Appendix A. Charged hadrons in $|\eta| < 2.5$ are used for the two particle correlation, and the p_T range of associated particles is $0.2 < p_T^{\text{Assoc}} < 3$ GeV/c. Events of $N^{\text{ch}} \geq 50$ corresponding to the highest 5% multiplicity are selected as high multiplicity events. The shape of short-range $\Delta\phi$ correlation functions in the lowest multiplicity range ($0 \leq N^{\text{ch}} < 10$) show a large p_T dependence in $0.2 < p_T^{\text{Trig}} < 2$ GeV/c, and the short-range $\Delta\phi$ correlation function of $0.2 < p_T^{\text{Trig}} < 0.5$ GeV/c in the high multiplicity bin also show a quite different shape compared to the $\Delta\phi$ correlation functions of other p_T^{Trig} bins.

The v_{22} from Fourier fits as well as v_{22}^{sub} are presented in Figure 8, and the left (right) panel shows the v_{22}^{sub} with *Method 1* (*Method 2*). In the non-flow subtraction results with *Method 1* using only long-range $\Delta\phi$ correlation functions, the v_{22}^{sub} strongly depends on p_T^{Trig} , and the deviation from zero becomes larger as p_T^{Trig} increases. The difference is largest when using events in the lowest multiplicity range ($0 \leq N^{\text{ch}} < 10$) as the reference. These v_{22}^{sub} results indicate that the shape of the long-range $\Delta\phi$ correlation function changes with multiplicity so that the scaled low multiplicity $\Delta\phi$ correlation does not perfectly describe those in high multiplicity events. The v_{22}^{sub} from *Method 2* in the right panel shows a smaller deviation from zero than that of *Method 1*, and the non-flow

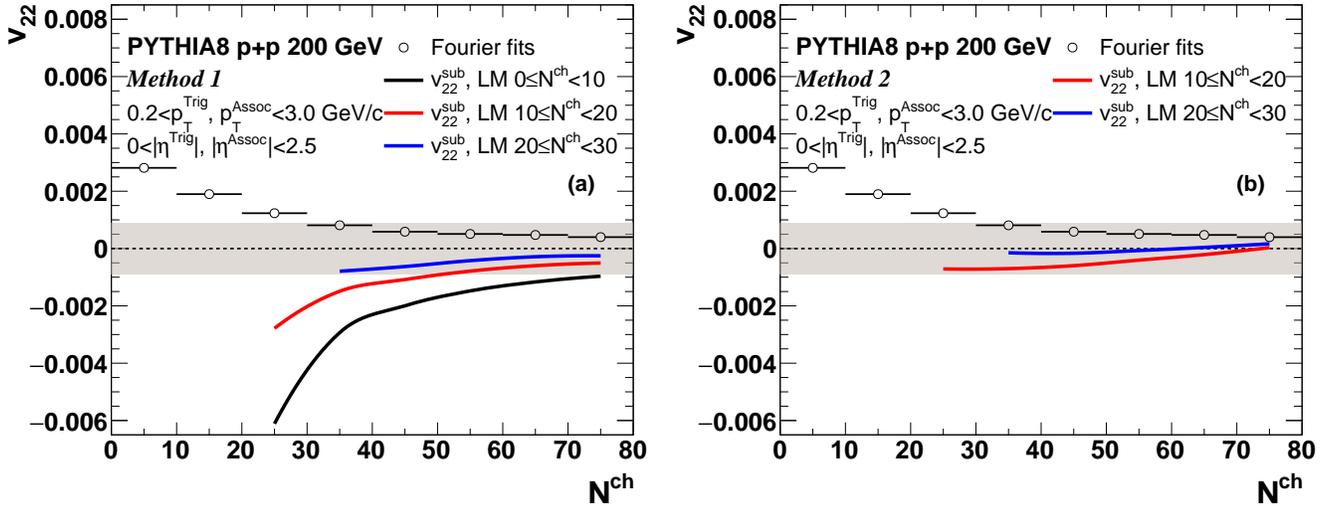


FIG. 7. The second order Fourier coefficient v_{22} of long-range ($2 < |\Delta\eta| < 5$) two particle correlation as a function of charged hadron multiplicity in $p+p$ collisions at $\sqrt{s} = 200$ GeV from PYTHIA8 before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/c and $|\eta| < 2.5$. Gray bands are corresponding to a 3% $|v_2|$ window.

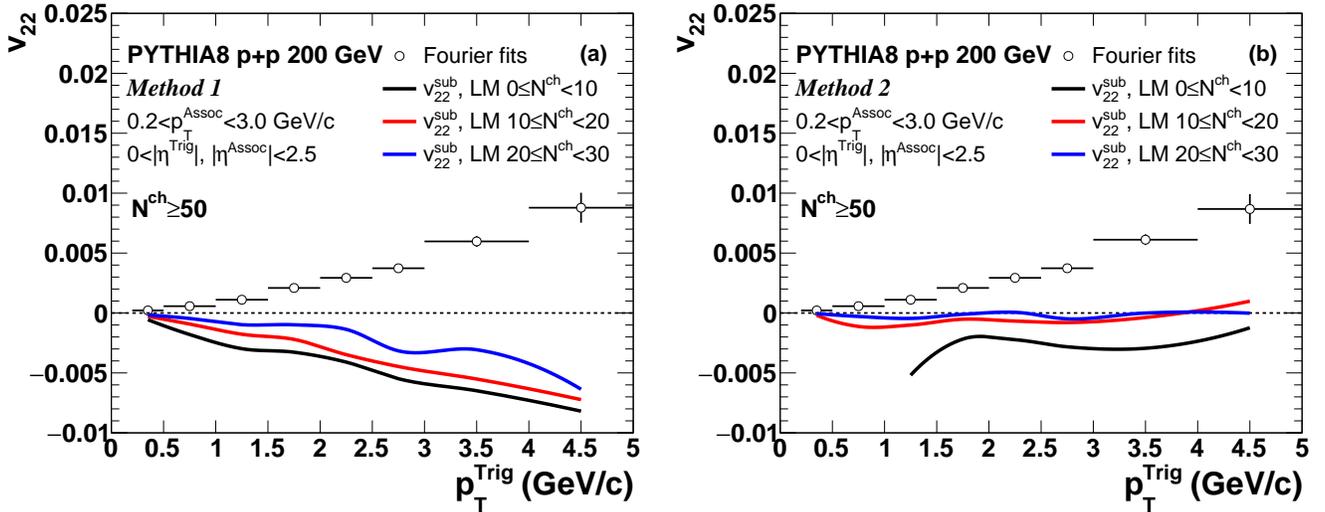


FIG. 8. The second order Fourier coefficient v_{22} of long-range ($2 < |\Delta\eta| < 5$) two particle correlation as a p_T in $p+p$ collisions of $N^{\text{ch}} \geq 50$ at $\sqrt{s} = 200$ GeV from PYTHIA8 before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/c and $|\eta| < 2.5$.

subtraction result using the lowest multiplicity reference ($0 \leq N^{\text{ch}} < 10$) is worse than the two other cases. It is notable that there is a much smaller shape variation when going from low to high multiplicity events when categorizing event activity at forward rapidity. This can be considered for future analyses with $p+p$ data at RHIC.

To summarize the closure test in $p+p$ collisions at $\sqrt{s} = 200$ GeV from PYTHIA8:

- The pseudorapidity coverage $|\eta| < 2.5$ was used to compare with the closure test results of the LHC (Section III A).
- The $\Delta\phi$ two particle correlation functions exhibit

a clear multiplicity and p_T^{Trig} dependent shape variation unlike the case in $p+p$ collisions at $\sqrt{s} = 13$ TeV from PYTHIA8.

- This shape variation results in a significant dependence on the low multiplicity reference selection for both methods.
- v_{22}^{sub} from *Method 1* shows a much larger deviation from zero compared to the results in the LHC case.
- v_{22}^{sub} from *Method 2* gives a smaller deviation than that of *Method 1*, but *Method 2* can not be applied using the some low multiplicity references and p_T^{Trig}

ranges due to a negative jet yield.

C. RHIC p +Au Case

There have been numerous extractions of flow coefficients at RHIC in p +Au, d +Au, and ^3He +Au collisions – highlighted by the PHENIX publications of elliptic v_2 and triangular v_3 flow in all three systems [6]. The PHENIX results are shown with no subtraction of the non-flow and instead with asymmetric systematic uncertainties to estimate the possible contributions. These correlations have been checked with a pseudorapidity gap as large as $|\Delta\eta| > 2.75$ using the PHENIX central arm tracks ($|\eta| < 0.35$) and the Au-going Beam-Beam Counter ($-3.9 < \eta < -3.1$).

Recently the STAR experiment has shown preliminary results using tracks in their Time Projection Chamber only, with pseudorapidity $|\eta| < 0.9$ and $|\Delta\eta| > 1.0$ [19]. The much smaller pseudorapidity gap yields a much larger non-flow contribution with influences on both the near-side $\Delta\phi \approx 0$ and the away-side $\Delta\phi \approx \pi$. They have employed multiple of the above outlined subtraction techniques to extract preliminary flow coefficients and find smaller v_2 than the PHENIX results particularly for $p_T > 1.5$ GeV/ c in high multiplicity p +Au and d +Au events [19].

Here we examine the non-flow subtraction in various kinematic ranges and multiplicity classifications in p +Au collisions. One billion p +Au and two billion p + p events are generated with HIJING [11], and charged hadrons are selected for two particle correlations. We explore a non-flow subtraction with p + p events from HIJING in addition to using low multiplicity or peripheral p +Au events. First, the case of a wide pseudorapidity coverage ($|\eta| < 2.5$) similar to the LHC experiments has been studied to check any difference with the same kinematic range but in lower collision energy (as was done with the study in PYTHIA8 p + p collisions at $\sqrt{s} = 200$ GeV). Then we detail a study modeling the more limited STAR acceptance.

Figure 22 in Appendix A shows two particle $\Delta\phi$ correlation function in short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) range in various multiplicity bins in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING. Charged hadrons in $0.2 < p_T < 3$ GeV/ c and $|\eta| < 2.5$ are used for the correlation functions, and the multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/ c and $|\eta| < 2.5$. Similar to the case of PYTHIA8 p + p in $\sqrt{s} = 200$ GeV, the shape of the two particle $\Delta\phi$ correlation function in the lowest multiplicity bin ($0 \leq N^{\text{ch}} < 10$) is quite different from that in the higher multiplicity ranges.

The v_{22} from direct Fourier fits to the two particle $\Delta\phi$ correlation in long-range are presented in Figure 9, and the lines represent the v_{22}^{sub} using low multiplicity events in three different ranges for non-flow subtraction. The v_{22}^{sub} with events in the lowest multiplicity range ($0 \leq N^{\text{ch}} < 10$) show the largest deviation from zero

in the results with both methods. The worse closure with *Method 2* is probably related to the shape of near-side short-range correlation in low multiplicity events ($0 \leq N^{\text{ch}} < 10$). The v_{22}^{sub} with events in the two other low multiplicity bins, $10 \leq N^{\text{ch}} < 20$ and $20 \leq N^{\text{ch}} < 30$, are within the window of $v_2 = 0.03$.

We further explore this closure test using kinematic ranges similar to the STAR experiment [19]. Figure 23 in Appendix A show the two particle $\Delta\phi$ correlation function in short ($0 < |\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) range in various multiplicity bins in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING. Charged hadrons in $0.2 < p_T < 3$ GeV/ c and $|\eta| < 0.9$ are used for the two particle correlations, and multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/ c and $|\eta| < 0.9$. In this narrower pseudorapidity acceptance, the shape of two particle $\Delta\phi$ correlation function in the lowest multiplicity bin ($0 \leq N^{\text{ch}} < 5$) is different from that in higher multiplicity ranges, as was the case with a wider pseudorapidity acceptance ($|\eta| < 2.5$). Another important thing to point out is that there is a clear peak shape at the near-side ($\Delta\phi \approx 0$) in long-range from jet correlations which is invisible in $\Delta\phi$ correlation functions with a larger $\Delta\eta$ gap ($2 < |\Delta\eta| < 5$) shown in Figure 22.

Figure 10 shows the v_{22} from Fourier fits to the two particle $\Delta\phi$ correlation in long-range ($1 < |\Delta\eta| < 1.8$), and the values are larger than those from the wider $\Delta\eta$ gap indicating a stronger non-flow effect with a smaller $\Delta\eta$ gap. The solid lines give v_{22}^{sub} using two different low multiplicity bins, $0 \leq N^{\text{ch}} < 5$ and $5 \leq N^{\text{ch}} < 10$. It is interesting that the v_{22}^{sub} using the lowest multiplicity bin is significantly different between the two methods. The positive v_{22}^{sub} with the *Method 1* is due to the remaining jet correlation with a smaller $\Delta\eta$ gap in $\Delta\phi$ correlation functions of higher multiplicity bins which is barely seen in the $\Delta\phi$ correlations of the lowest multiplicity. Therefore, the scaled correlation function of the low multiplicity bin cannot describe the peak structure on the near-side, resulting in the positive v_{22}^{sub} . The results from *Method 2* show large negative values, and this is related to the different shape of short-range $\Delta\phi$ correlation function in the lowest multiplicity bin introducing a large scaling with jet yields. The v_{22}^{sub} with the next low multiplicity bin ($5 \leq N^{\text{ch}} < 10$) show a better closure within the level of $|v_2| = 0.03$.

Another way to categorize event activity in p +Au collisions is to use centrality defined with charged particle multiplicity in the Au-ion-going (backward) rapidity ($-5.0 < \eta < -3.3$ in STAR and $-3.9 < \eta < -3.1$ in PHENIX). Results from the PHENIX and STAR experiments in small systems are categorized in this manner. In p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV, the multiplicity correlation between mid and backward rapidity is weak, so the shape difference of the $\Delta\phi$ correlation functions seen in low multiplicity events defined at midrapidity may not appear in peripheral events defined at backward rapidity. Figure 24 in Appendix A shows two particle $\Delta\phi$ correlation in short ($|\Delta\eta| < 0.5$) and long

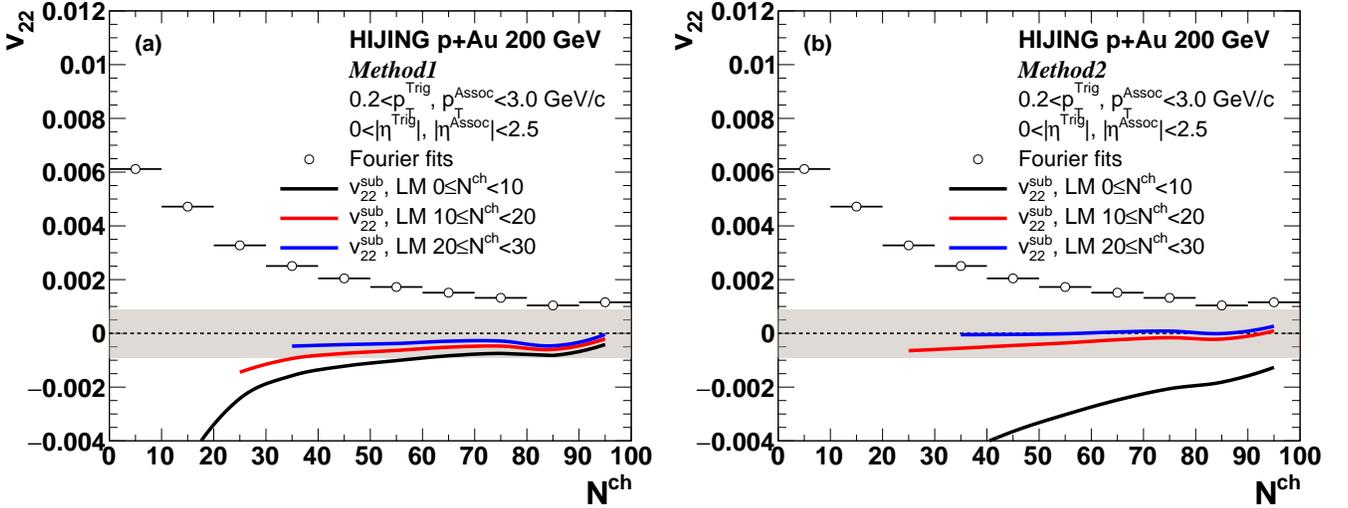


FIG. 9. The second order Fourier coefficient v_{22} of long-range ($2 < |\Delta\eta| < 5$) two particle correlation as a function of charged hadron multiplicity in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/c and $|\eta| < 2.5$. Gray bands correspond to a 3% $|v_2|$ window.

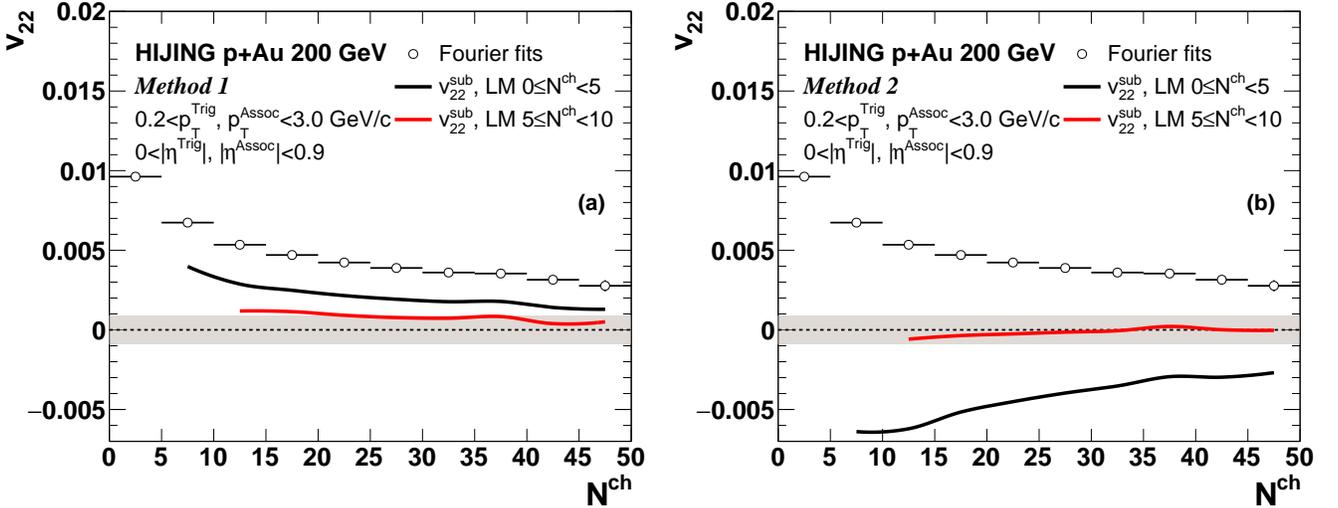


FIG. 10. The second order Fourier coefficient v_{22} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of charged hadron multiplicity in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/c and $|\eta| < 0.9$. Gray bands are corresponding to a 3% $|v_2|$ window.

($1 < |\Delta\eta| < 1.8$) ranges in p + p and various centrality bins of p +Au collisions from HIJING. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 0.9$ are used for the correlation functions, and centrality is defined with charged hadrons in $-5.0 < \eta < -3.3$. As discussed earlier, one thing notably different from the case of multiplicity categorized at midrapidity shown in Figure 23 is that the shape of two particle $\Delta\phi$ correlation function is similar in p + p and all centrality bins of p +Au collisions. At RHIC, selecting events based on multiplicity in the same kinematic range as the particles selected for correlations introduces a significant undesirable shape variation of

correlation functions.

Figure 11 shows the v_{22} from Fourier fits to the two particle $\Delta\phi$ correlation at long-range. Here, the non-flow effects become larger from peripheral events to central events. The lines are v_{22}^{sub} using events from p + p and two different peripheral selections (50–85% and 85–100%) of p +Au events. Note that it is usually difficult to collect events of 85–100% centrality of p +Au collisions in experiments due to the difficulty of triggering on events with small multiplicity at forward and backward rapidity, so the next peripheral bin (50–85%) is also used for the non-flow subtraction. The v_{22}^{sub} from both methods with all

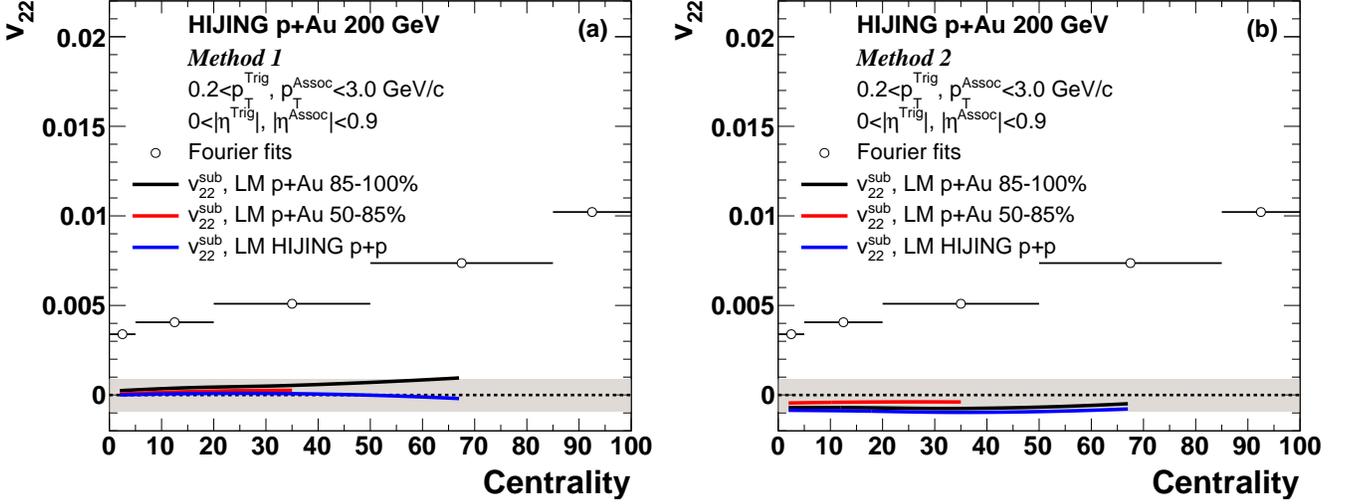


FIG. 11. The second order Fourier coefficient v_{22} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of centrality in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING before and after non-flow subtraction. Centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction). Gray bands correspond to a 3% $|v_2|$ window.

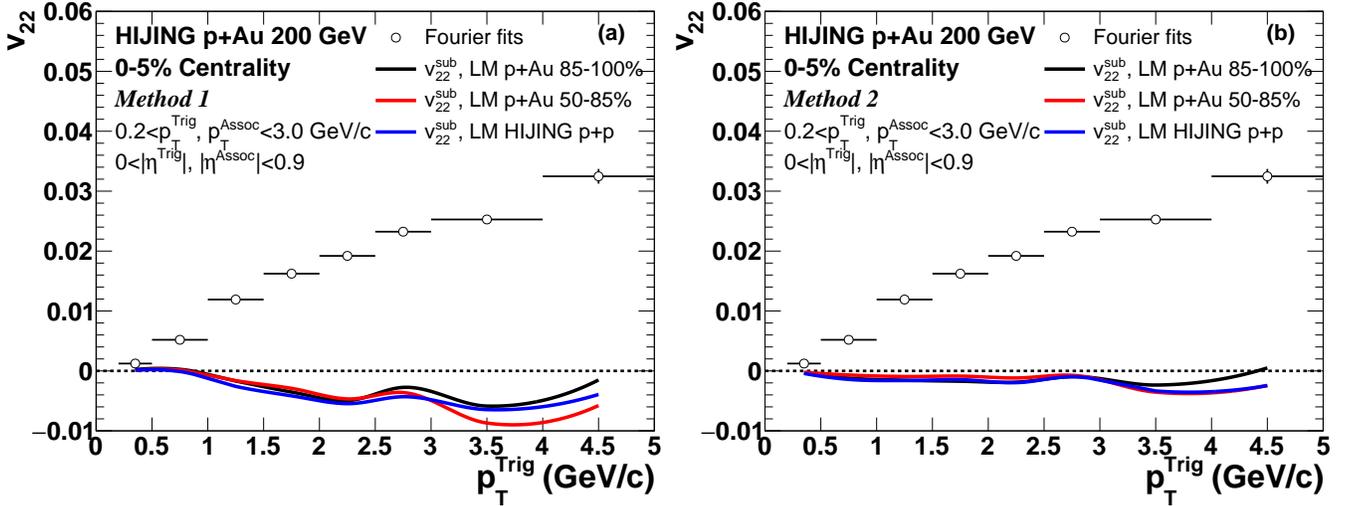


FIG. 12. The second order Fourier coefficient v_{22} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of p_T in 0-5% of p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING before and after non-flow subtraction. Centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction).

three selections of low multiplicity events are within the level of $|v_2| = 0.03$ shown as a gray band. This smaller v_{22}^{sub} with centrality compared to the results using multiplicity at midrapidity is mainly coming from the similar shape of the $\Delta\phi$ correlation functions from peripheral to central p +Au collisions.

As an additional test in p +Au collisions from HIJING, we have checked the p_T^{Trig} dependence in 0-5% central collisions. Figure 25 in Appendix A shows $\Delta\phi$ correlations in short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) range for different p_T^{Trig} bins. Charged hadrons in $|\eta| < 0.9$ are used for the two particle correlation, and the p_T range of associated particles is $0.2 < p_T^{\text{Assoc}} < 3$ GeV/c. Central-

ity is defined with charged hadrons in $-5.0 < \eta < -3.3$ (Au-going direction). The shape of the correlation function in $0.2 < p_T^{\text{Trig}} < 0.5$ GeV/c is different from other p_T^{Trig} bins both in p +p and 0-5% p +Au collisions, but the shapes in p +p and 0-5% p +Au events at the same p_T^{Trig} bin look comparable.

Figure 12 shows the v_{22} from Fourier fit to the long-range $\Delta\phi$ correlations as a function of p_T^{Trig} in 0-5% central p +Au collisions, and the lines represent v_{22}^{sub} with events in p +p and two peripheral centrality bins of p +Au collisions. Note that the y -axis range is much larger than the previous plots where the p_T -integrated results are

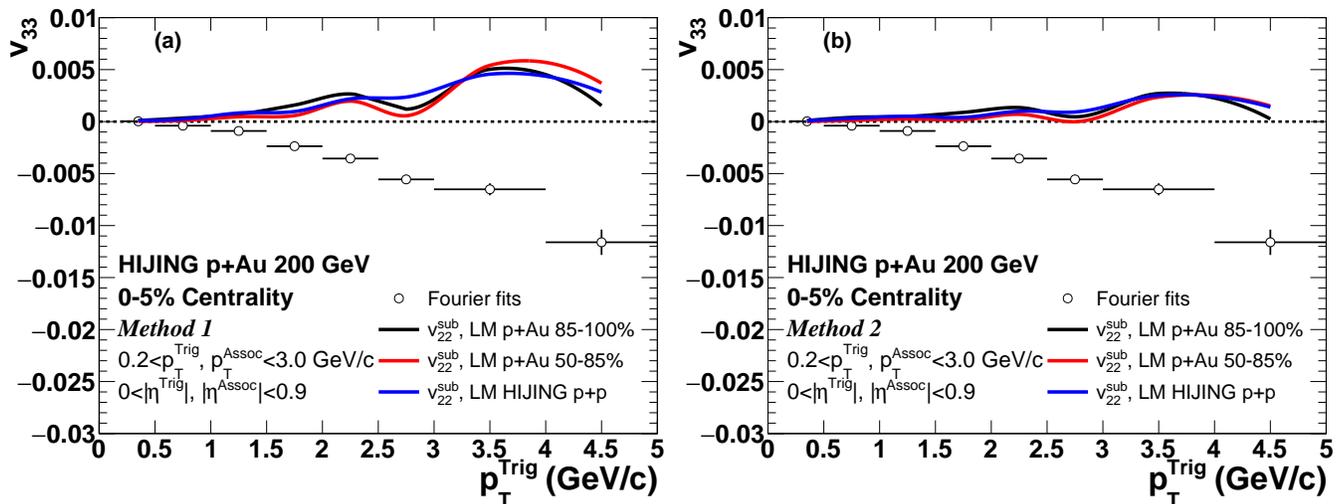


FIG. 13. The third order Fourier coefficient v_{33} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of p_T in 0–5% of p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING before and after non-flow subtraction. Centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction).

shown as a function of multiplicity and centrality, due to a much stronger non-flow contribution in the higher p_T^{Trig} range than that in the p_T -integrated case. v_{22}^{sub} shows a weak dependence on the selection of low multiplicity events for both methods. However, the results from *Method 1* show a clear p_T^{Trig} dependence and large negative values indicating a significant over-subtraction similar to the case of p + p collisions at $\sqrt{s} = 200$ GeV shown in Figure 8. *Method 2* shows a smaller deviation from zero than *Method 1*, but the v_{22}^{sub} values are still larger than 0.001.

Figure 13 shows the v_{33} from the Fourier fit and the subtraction methods. Unlike the v_{22} case, v_{33} from the Fourier fit yields negative values. With the subtraction procedures applied, v_{33}^{sub} from both methods yields positive values. Thus the methods result in an over subtraction, but with a negative sign, hence leading to a strong p_T -dependent enhancement.

The PHENIX results are measured via multiple detector systems covering a wide range in pseudorapidity. However, the midrapidity coverage is limited ($|\eta| < 0.35$) and thus the short-range correlations used in *Method 2* are not available. The PHENIX flow measurements generally use the event plane method, but there are comparisons using three sets of two-particle correlations, where one can then algebraically solve for the anisotropy at midrapidity. In principle one can apply *Method 1* to each of the three sets of two-particle correlations. We reserve this more detailed study to a future publication.

To summarize the closure test in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING:

- There are very substantial jet shape modifications when selecting events based on charged particle multiplicity around midrapidity N^{ch} . These lead to significant distortions in the non-flow subtrac-

tion in both *Method 1* and *2* and for a larger and smaller acceptance.

- Selection of event categories based on forward detectors away from midrapidity, as employed by PHENIX and STAR, significantly improve the results of the HIJING closure test.
- The v_{22}^{sub} in integrated p_T as a function of centrality with the two methods have some dependence on the low multiplicity reference selection, but are generally within the $|v_2| < 0.03$ level in both *Method 1* and *2*.
- In contrast the p_T dependent results indicate a significant over-subtraction of non-flow in both *Method 1* and *2*. The over-subtraction is very large in *Method 1* for $p_T > 1$ GeV/ c .
- In case of the v_{33}^{sub} as a function of p_T , there is an indication of significant bias to increase v_{33}^{sub} in both *Method 1* and *2*.

IV. FURTHER TESTS WITH AMPT

The above tests have a significant limitation in that one is testing procedures to disentangle flow and non-flow on models that have only non-flow. AMPT [12] is a model that includes both contributions and can thus be further elucidating. We follow the previous AMPT studies in Refs. [20] to obtain the “truth” flow with respect to the participant plane calculated by using initial-state coordinates of partons resulting from string melting. More details on the method to calculate the truth flow can be found in Ref. [20]. We highlight that the finite number of

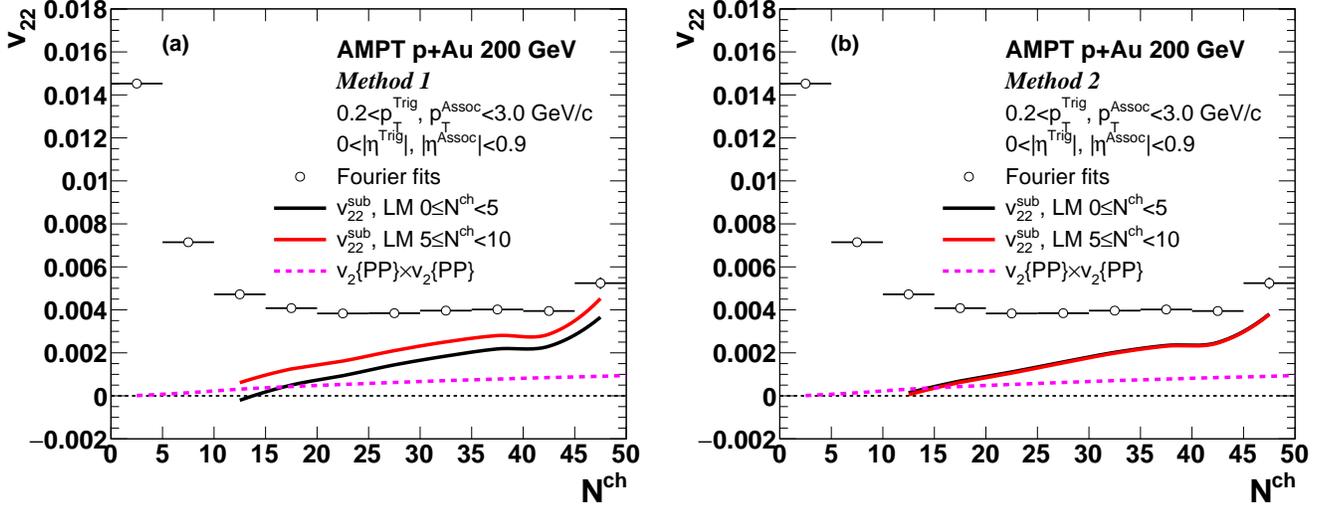


FIG. 14. The second order Fourier coefficient v_{22} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of charged particle multiplicity in $p+Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT before and after non-flow subtraction. Multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/ c and $|\eta| < 0.9$.

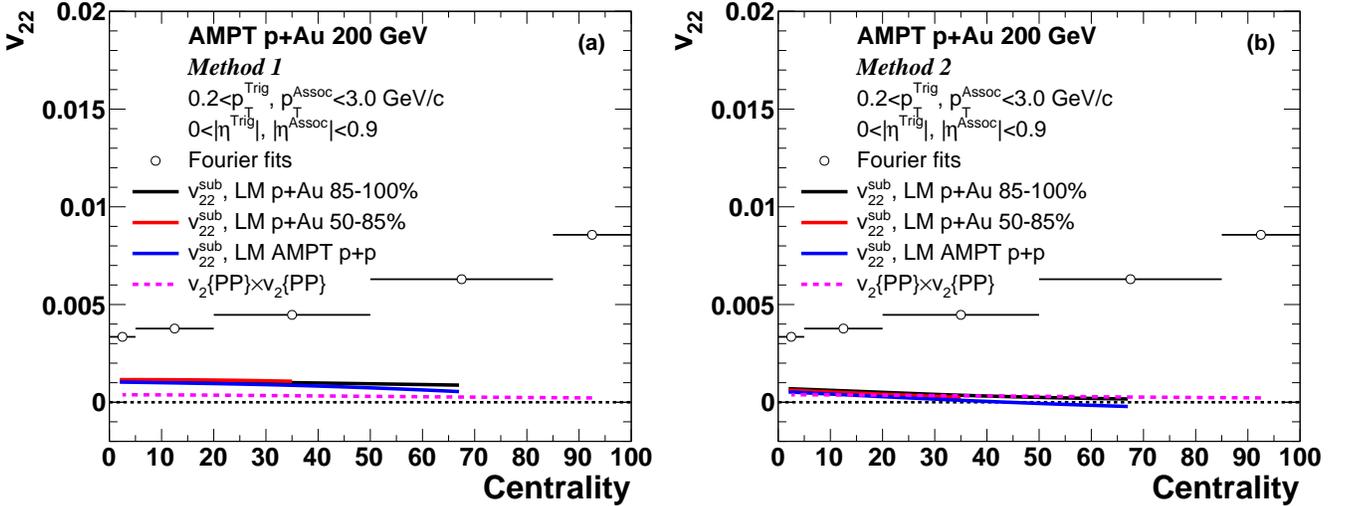


FIG. 15. The second order Fourier coefficient v_{22} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of centrality in $p+Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT before and after non-flow subtraction. Centrality is defined as the number of charged hadrons in $-5.0 < \eta < -3.3$ (Au-going direction).

partons used to define the geometry and possible factorization breaking make this a rough estimate of the truth. A key item to note is that previous AMPT studies [21] indicate that flow and non-flow do not factorize, in part because partons from jets can rescatter with medium partons. For both of these reasons, we are not expecting a perfect closure but rather examining possible trends in the subtraction procedure.

For this test, we used AMPT v2.26 with string melting, and a few important parameters are listed in Table I. Note that the updated string parameters, PARJ(41) and PARJ(42), introduced in Ref. [22] have been found to be important in recovering the peak structure from jets in

near-side short-range correlation functions.

TABLE I. Parameters used in AMPT.

Parameter	Value
ISOFT	4
PARJ(41)	0.55
PARJ(42)	0.15
Parton screening mass	6.45d0 (0.75 mb)
α in parton cascade	0.47d0

We follow the same procedure applied above with

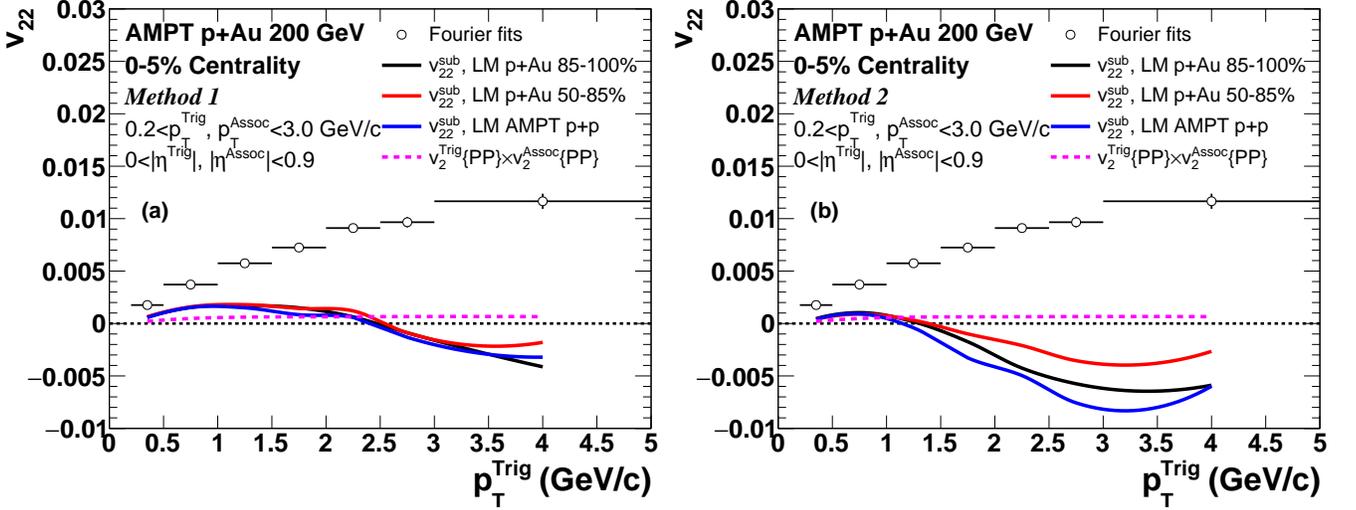


FIG. 16. The second order Fourier coefficient v_{22} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of p_T in 0–5% of p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT before and after non-flow subtraction. Centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction).

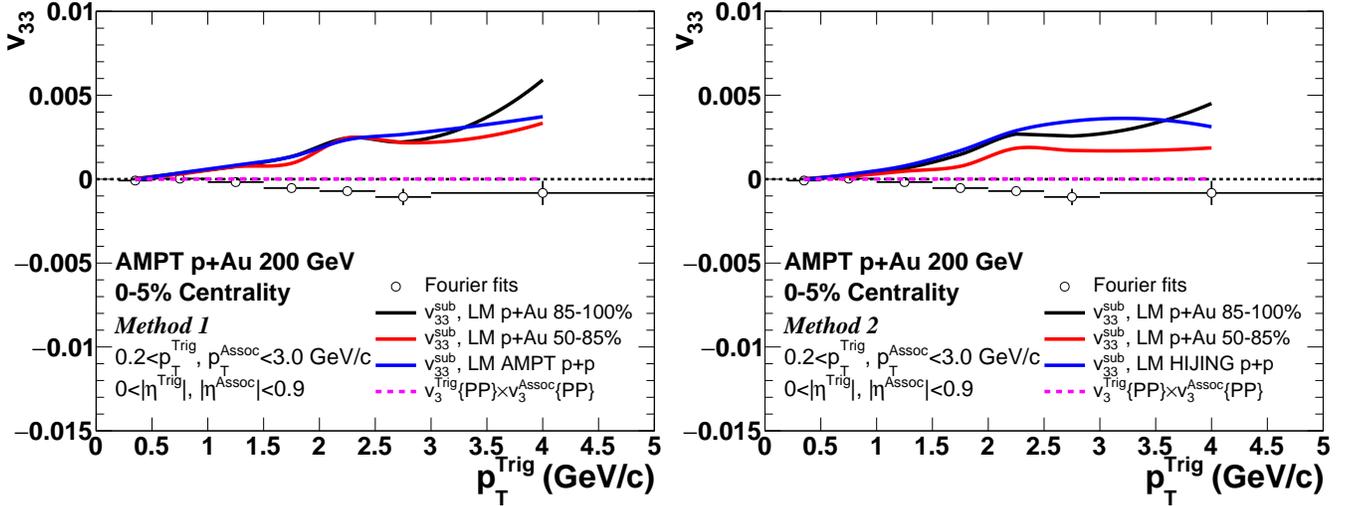


FIG. 17. The third order Fourier coefficient v_{33} of long-range ($1 < |\Delta\eta| < 1.8$) two particle correlation as a function of p_T in 0–5% of p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT before and after non-flow subtraction. Centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction).

HIJING p +Au and p + p events. Two particle $\Delta\phi$ and $\Delta\eta$ correlations are made from charged hadrons from 600 million p +Au and 200 million p + p AMPT events. Figure 26 in Appendix A shows the two-dimensional correlation functions in short ($0 < |\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) range and various multiplicity bins in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 0.9$ are used for the $\Delta\phi$ correlation functions, and multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/c and $|\eta| < 0.9$. One thing to point out from the comparison of correlations between AMPT and HIJING (shown in Figure 23) is that there is a more pro-

nounced peak structure at the near-side ($\Delta\phi \approx 0$) in the long-range correlations which has contributions from the truth flow in AMPT. Another difference is that the shape of the short-range $\Delta\phi$ correlation function in the lowest multiplicity bin ($0 \leq N^{\text{ch}} < 5$) is similar to the shape in higher multiplicity bins in AMPT events so that the non-flow subtraction results may not depend as much on the selection of the low multiplicity bin.

Figure 14 shows the v_{22} from Fourier fits to the two particle $\Delta\phi$ correlation functions in long-range as a function of multiplicity at midrapidity. As with the previous studies, the solid lines are the v_{22}^{sub} with two different selections of low multiplicity events, and the dashed line

is v_{22} calculated as a product of truth flow ($v_2^2\{\text{PP}\}$) with an assumption of factorization. The v_{22}^{sub} with both methods and $v_2^2\{\text{PP}\}$ are close to zero for low multiplicity events and increase with multiplicity, but the v_{22}^{sub} is much larger than $v_2^2\{\text{PP}\}$. One caveat is v_{22} , which is a root-mean-square of v_2 ($\langle v_2^2 \rangle$), is higher than a product of average v_2 ($\langle v_2 \rangle^2$). We have estimated the difference via the relation $\langle v_2^2 \rangle / \langle v_2 \rangle^2 \propto \langle \varepsilon_2^2 \rangle / \langle \varepsilon_2 \rangle^2$ is $\sim 20\%$, which is much smaller than the observed difference.

Another set of two particle $\Delta\phi$ correlations in short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) ranges with charged hadrons in the same kinematic range but different event multiplicity categorization defined at backward rapidity are shown in Figure 27 in Appendix A. From $p+p$ collisions to various centrality ranges of $p+Au$ collisions, the shape of the correlations are quite similar except for the near-side peak structure at long-range, possibly due to the truth flow in AMPT. Figure 15 shows the v_{22} from Fourier fits in long-range as a function of centrality. The solid lines are the v_{22}^{sub} with two different selections of low multiplicity events, and the dashed lines represent v_{22} from the truth flow, $v_2^2\{\text{PP}\}$. The v_{22}^{sub} with the *Method 1* in central $p+Au$ collisions is larger than the $v_2^2\{\text{PP}\}$ like the v_{22}^{sub} as a function of multiplicity at midrapidity. The v_{22}^{sub} using *Method 2* is smaller than the v_{22}^{sub} from *Method 1* and is consistent with the $v_2^2\{\text{PP}\}$.

An additional test with AMPT events is done for different p_T^{Trig} bins in 0–5% central $p+Au$ collisions. Figure 28 in Appendix A shows the two particle $\Delta\phi$ correlations in short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) range for various p_T^{Trig} bins in $p+p$ and 0–5% central $p+Au$ collisions from AMPT. The correlations are made from charged hadrons in $|\eta| < 0.9$, and the p_T range of associated particles is $0.2 < p_T^{\text{Assoc}} < 3$ GeV/ c . Centrality is defined with charged hadron multiplicity in $-5.0 < \eta < -3.3$ in the direction of the Au-ion. Like the case of HIJING $p+p$ and $p+Au$ events, the shape of the two particle $\Delta\phi$ correlations in the lowest p_T^{Trig} bin is different from that in other p_T^{Trig} bins both in $p+p$ and 0–5% $p+Au$ collisions, but the shape at the same p_T^{Trig} bin of $p+p$ events and 0–5% $p+Au$ events looks comparable.

Figure 16 shows the v_{22} from Fourier fits to the long-range two particle $\Delta\phi$ correlation as a function of p_T^{Trig} in 0–5% central $p+Au$ collisions. The solid lines are v_{22}^{sub} with low multiplicity selections from $p+p$ and two peripheral $p+Au$ centrality bins, and the dashed line represents the v_{22} from the truth flow, $v_2^{\text{Trig}}\{\text{PP}\} \times v_2^{\text{Assoc}}\{\text{PP}\}$. Similar to the case of HIJING $p+Au$ events shown in Figure 12, the non-flow effects increase with p_T^{Trig} . The v_{22}^{sub} with both methods are below zero in $p_T^{\text{Trig}} > 2.5$ GeV/ c using *Method 1* and $p_T^{\text{Trig}} > 1$ GeV/ c using *Method 2*. This clearly indicates an over-subtraction in these p_T^{Trig} ranges, because the v_{22} from truth flow in AMPT should give positive values even though the truth v_{22} from two particle correlation may not be exactly same as

$v_2^{\text{Trig}}\{\text{PP}\} \times v_2^{\text{Assoc}}\{\text{PP}\}$ shown as the dashed line. The deviation of the v_{22}^{sub} from zero is larger with *Method 2* which is opposite to the closure test results with HIJING events shown in Figure 12. This is possibly due to the different shape of two particle $\Delta\phi$ correlation function between HIJING and AMPT both in short and long ranges presented in Figures 25 and 28.

Figure 17 shows the v_{33} from the Fourier fit and the subtraction methods. Note that the v_{33} from the truth flow, $v_3^{\text{Trig}}\{\text{PP}\} \times v_3^{\text{Assoc}}\{\text{PP}\}$, are non-zero but have very small values. Similar to the HIJING study, v_{33} from the Fourier fit yields negative values, but the magnitude is much smaller in AMPT, possibly due to the truth flow contribution in AMPT. In the subtraction results, v_{33}^{sub} from both methods yields positive values that are significantly larger than the truth flow. As was the case in the HIJING study, there is an over-subtraction, again with a negative sign, and thus a strong p_T -dependent enhancement.

The summary of the AMPT studies is as follows:

- Selection of $p+Au$ event categories based on midrapidity multiplicity N^{ch} result in significant under-subtraction of non-flow contributions in AMPT.
- Selection of $p+Au$ event categories based on forward detectors away from midrapidity, as employed by PHENIX and STAR, improve the results of the AMPT test. The results as a function of centrality for the v_{22}^{sub} values in the two methods match the truth extracted AMPT flow qualitatively.
- In contrast the p_T dependent results indicate a significant over-subtraction of non-flow in both *Method 1* and *2*. The over-subtraction is very large in *Method 2* for $p_T > 1$ GeV/ c .
- In case of the v_{33}^{sub} as a function of p_T , there is an indication of significant bias to increase v_{33}^{sub} in both *Method 1* and *2*.

V. CONCLUSIONS

We examined closure tests of the non-flow subtraction methods used with two particle correlations developed by ATLAS and CMS Collaborations to study the collective behavior of particle production in small collision systems at the LHC. Monte Carlo event generators (PYTHIA and HIJING) including no collective flow are used to quantify a level of closure. In the test for $p+p$ collisions at $\sqrt{s} = 13$ TeV at the LHC where a large pseudorapidity gap ($|\Delta\eta| > 2$) can be applied, the shape of the $\Delta\phi$ correlation functions in short and long range is relatively stable throughout the event multiplicity and p_T ranges for which the resulting second-order Fourier coefficients after non-flow subtraction (v_{22}^{sub}) are less than 0.001 in these ranges.

We also tested the non-flow subtraction methods for $p+p$ collisions at $\sqrt{s} = 200$ GeV, and it is observed that

the shape of the $\Delta\phi$ correlation functions changes significantly with event multiplicity and p_T ranges. Therefore, the non-flow subtraction results are very sensitive to the selection of the low multiplicity reference used for the subtraction procedure, indicating that one should be extremely careful applying the same non-flow subtraction techniques to the $p+p$ data at RHIC in future analyses.

In the test with HIJING for $p+Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV, the large sensitivity to the low multiplicity event selection that was observed in $p+p$ collisions at $\sqrt{s} = 200$ GeV is still apparent when event activity is categorized at the same rapidity range as that of the two particle correlations. The test with a smaller $\Delta\eta$ gap ($1 < |\Delta\eta| < 1.8$) shows a significant difference between the methods, because it starts to be affected by short-range jet correlation contributions. Although the event activity categorization at backward rapidity helps to reduce the dependence of the low multiplicity event selection on the results of integrated p_T ($0.2 < p_T < 3$ GeV/ c), the test results as a function of p_T show a clear over-subtraction in $p_T > 1$ GeV/ c with both methods for v_2 and v_3 .

We extended this study with AMPT, which includes both flow and non-flow effects, and the non-flow subtraction results were compared to the truth flow with respect to the participant plane. The v_{22}^{sub} and v_{33}^{sub} are inconsistent with the truth flow in AMPT, and the p_T dependent results show negative values at higher p_T with both methods. Thus in $p+Au$ collisions, both HIJING and AMPT studies indicate that both *Method 1* and *Method 2* result in an over-subtraction of non-flow and thus an underestimate (overestimate) of the real v_2 (v_3).

ACKNOWLEDGMENTS

We acknowledge Jiangyong Jia and Wei Li for useful discussions and a careful reading of the manuscript. We acknowledge Julia Velkovska for useful discussions and the suggestion to extend this study to third order flow coefficients. SHL, KKH, JLN acknowledge support from the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-FG02-00ER41152. QH and DVP acknowledge support from the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-SC0018117.

Appendix A: Two particle $\Delta\phi$ correlation functions

Here we present two particle $\Delta\phi$ correlation functions described as Eq. 1 in Section I with charged hadrons from event generators, and these correlation functions are used for v_{22} and v_{22}^{sub} calculations. We specifically follow the formula for per-trigger yields as a function of $\Delta\phi$ within

a certain $\Delta\eta$ range described in Ref. [16],

$$\begin{aligned} f(\Delta\phi) &= \int \frac{1}{N^{\text{Trig}}} \frac{d^2 N^{\text{Pair}}}{d\Delta\eta d\Delta\phi} d\Delta\eta \\ &= \int B(0, 0) \frac{S(\Delta\eta, \Delta\phi)}{B(\Delta\eta, \Delta\phi)} d\Delta\eta, \end{aligned} \quad (\text{A1})$$

where $\Delta\eta$ and $\Delta\phi$ are the differences of η and ϕ between trigger and associated particles. $S(\Delta\eta, \Delta\phi)$ and $B(\Delta\eta, \Delta\phi)$ are the yields of pairs normalized by the number of trigger particles, N^{Trig} , in the same and mixed events respectively,

$$S(\Delta\eta, \Delta\phi) = \frac{1}{N^{\text{Trig}}} \frac{d^2 N^{\text{Same}}}{d\Delta\eta d\Delta\phi}, \quad (\text{A2})$$

$$B(\Delta\eta, \Delta\phi) = \frac{1}{N^{\text{Trig}}} \frac{d^2 N^{\text{Mix}}}{d\Delta\eta d\Delta\phi}. \quad (\text{A3})$$

The purpose of normalizing by the mixed event pair distributions is to account for the geometric $\Delta\eta$ -dependent pair acceptance effect.

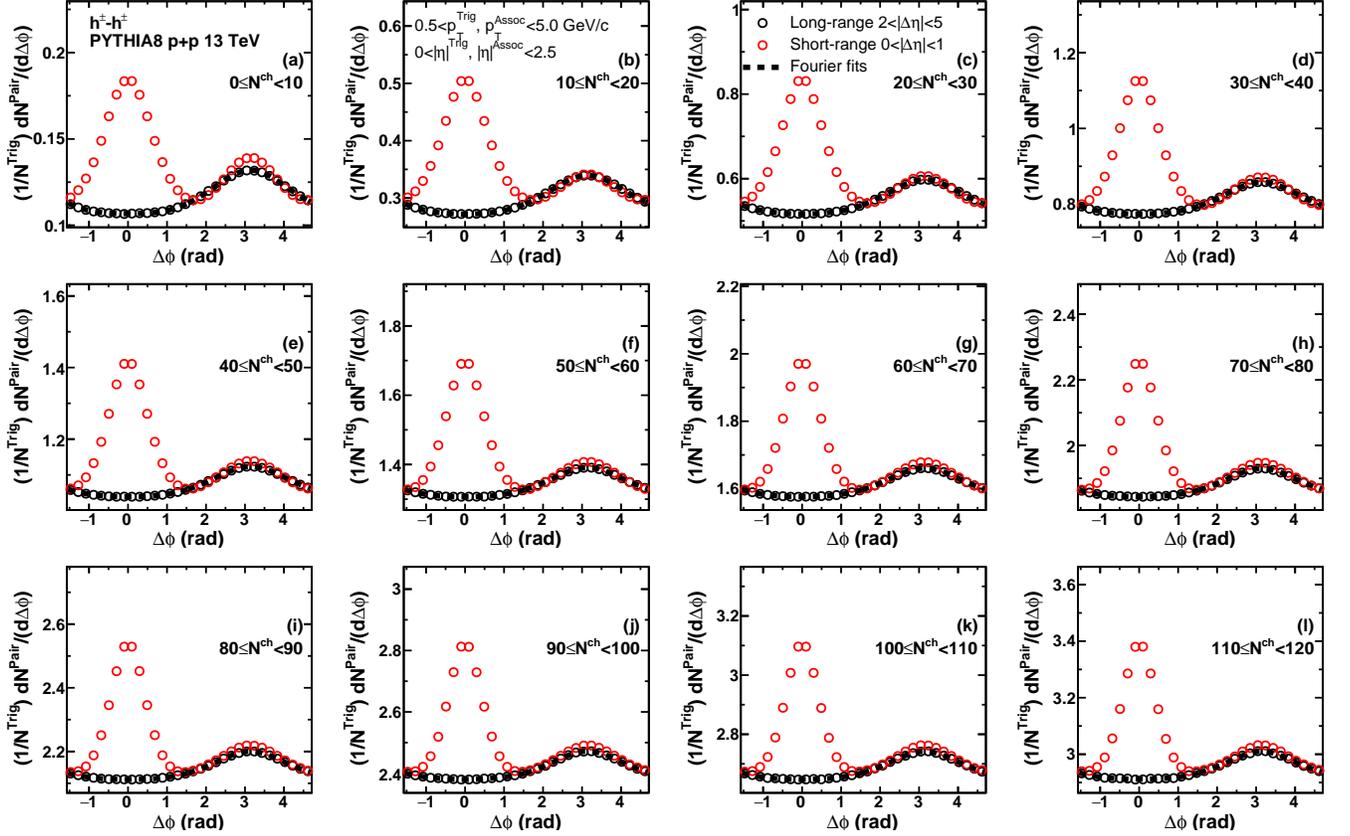


FIG. 18. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) ranges in $p+p$ collisions at $\sqrt{s} = 13$ TeV from PYTHIA8. Charged hadrons in $0.5 < p_T < 5$ GeV/c and $|\eta| < 2.5$ are used for the correlation function. Each panel show a different multiplicity range, and the multiplicity is defined as the number of charged hadrons in $p_T > 0.4$ GeV/c and $|\eta| < 2.5$.

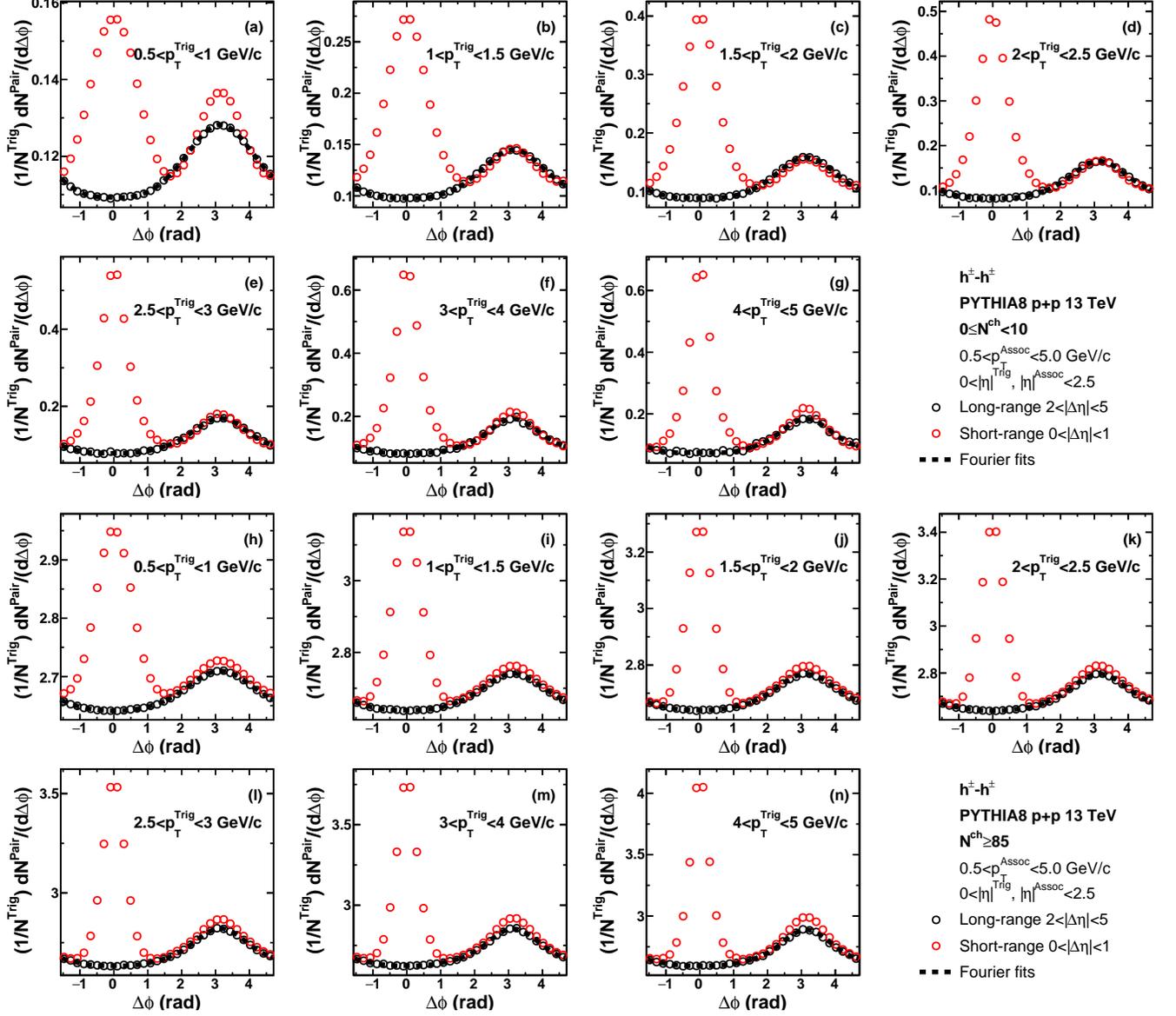


FIG. 19. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) ranges in $p+p$ collisions at $\sqrt{s} = 13$ TeV from PYTHIA8. Charged hadrons in $0.5 < p_T < 5$ GeV/ c and $|\eta| < 2.5$ are used for the correlation function. Panels in top (a)–(g) (bottom (h)–(n)) two rows are in a multiplicity range of $0 \leq N^{\text{ch}} < 10$ ($N^{\text{ch}} \geq 85$), and the multiplicity is defined as the number of charged particles in $p_T > 0.4$ GeV/ c and $|\eta| < 2.5$. Each panel represent a different p_T range of trigger particles.

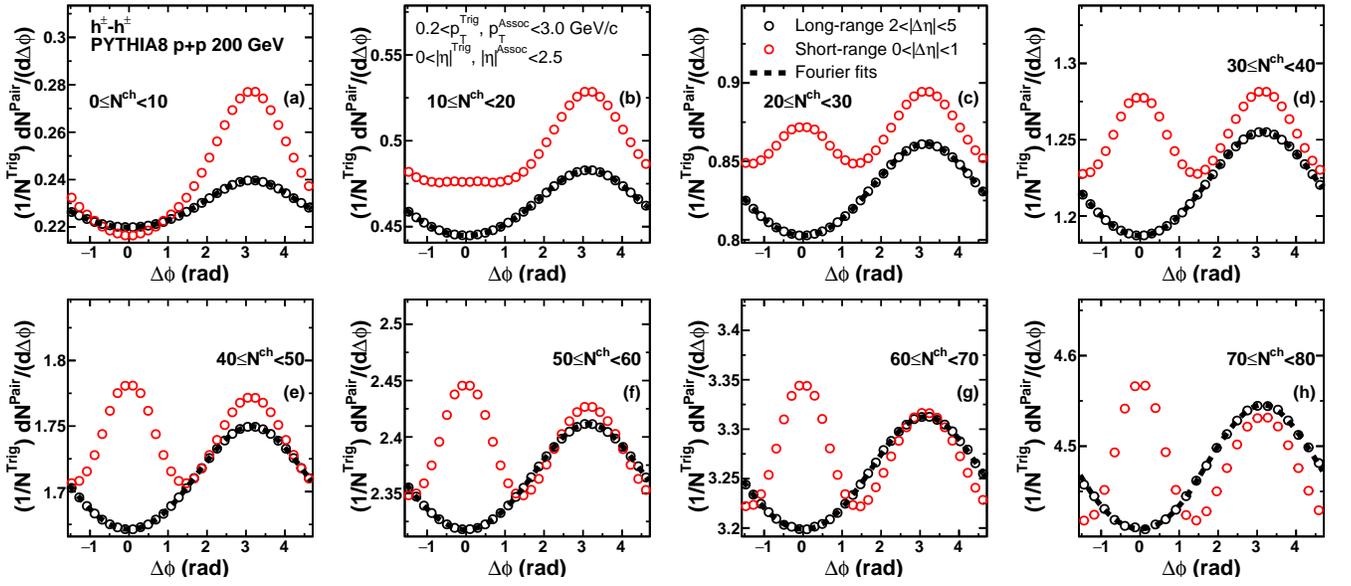


FIG. 20. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) ranges in $p+p$ collisions at $\sqrt{s} = 200$ GeV from PYTHIA8. Charged hadrons in $0.2 < p_T < 3$ GeV/ c and $|\eta| < 2.5$ are used for the correlation function. Each panel show a different multiplicity range, and the multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/ c and $|\eta| < 2.5$.

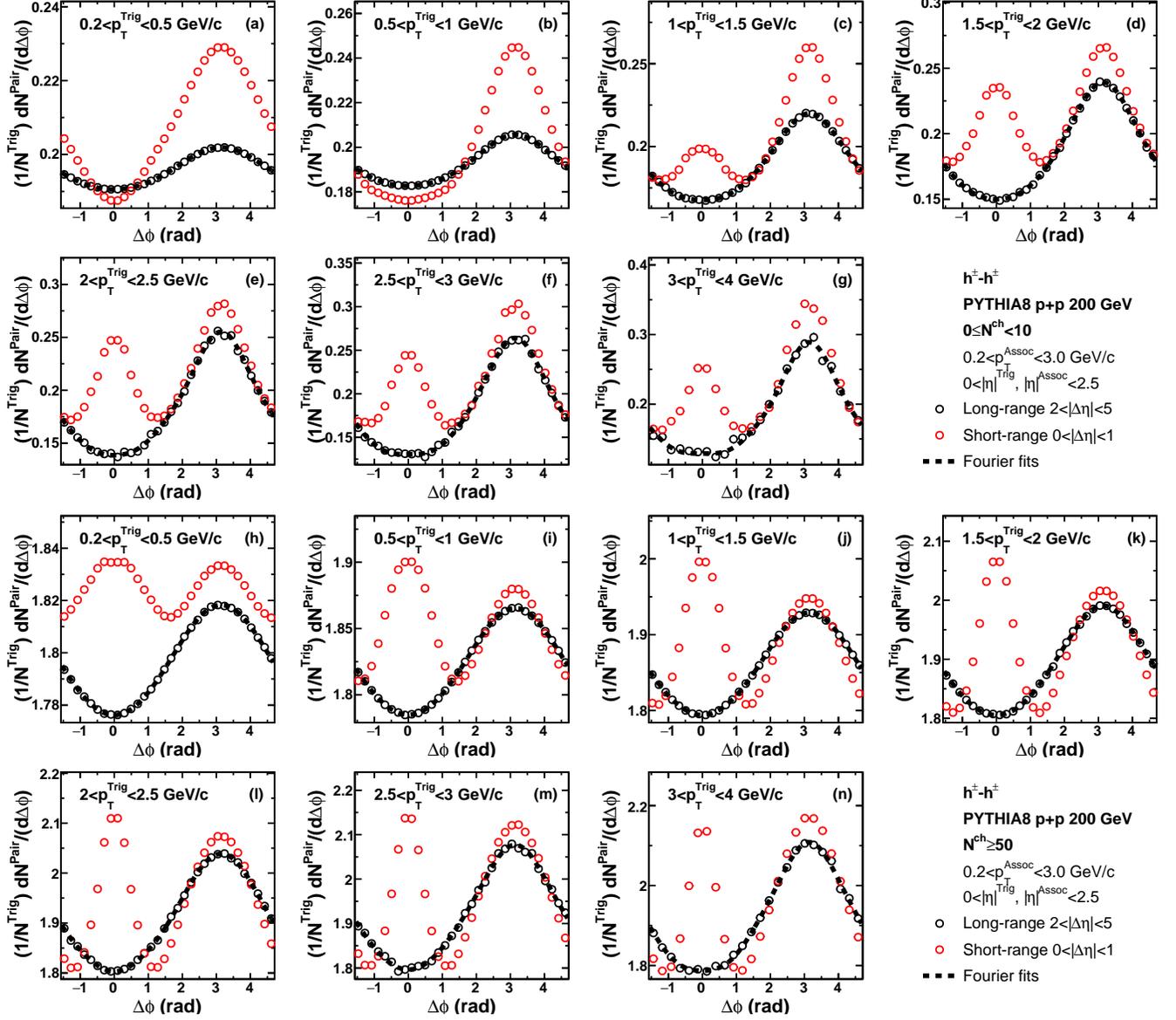


FIG. 21. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) ranges in $p+p$ collisions at $\sqrt{s} = 200$ GeV from PYTHIA8. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 2.5$ are used for the correlation function. Panels in top (a)–(g) (bottom (h)–(n)) two rows are in a multiplicity range of $0 \leq N^{\text{ch}} < 10$ ($N^{\text{ch}} \geq 50$), and the multiplicity is defined as the number of charged hadrons in $p_T > 0.2$ GeV/c and $|\eta| < 2.5$. Each panel represent a different p_T range of trigger particles.

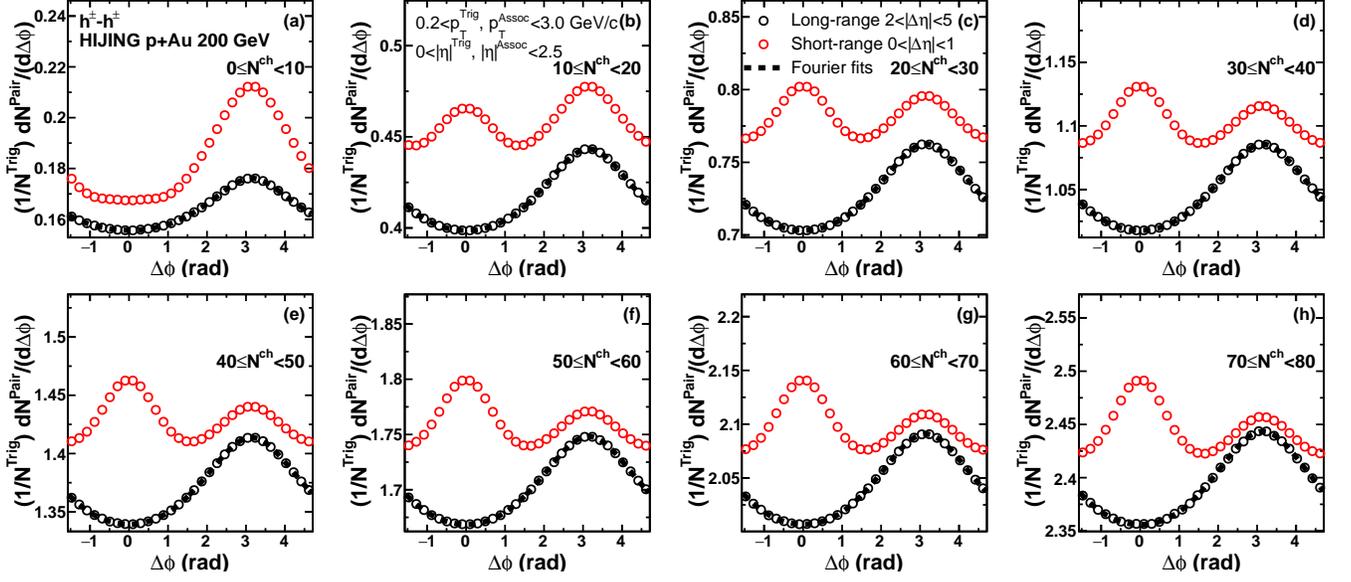


FIG. 22. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 1$) and long ($2 < |\Delta\eta| < 5$) ranges in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING. Charged hadrons in $0.2 < p_T < 3$ GeV/ c and $|\eta| < 2.5$ are used for the correlation function. Each panel show a different multiplicity range, and the multiplicity is defined as the number of charged particles in $p_T > 0.2$ GeV/ c and $|\eta| < 2.5$.

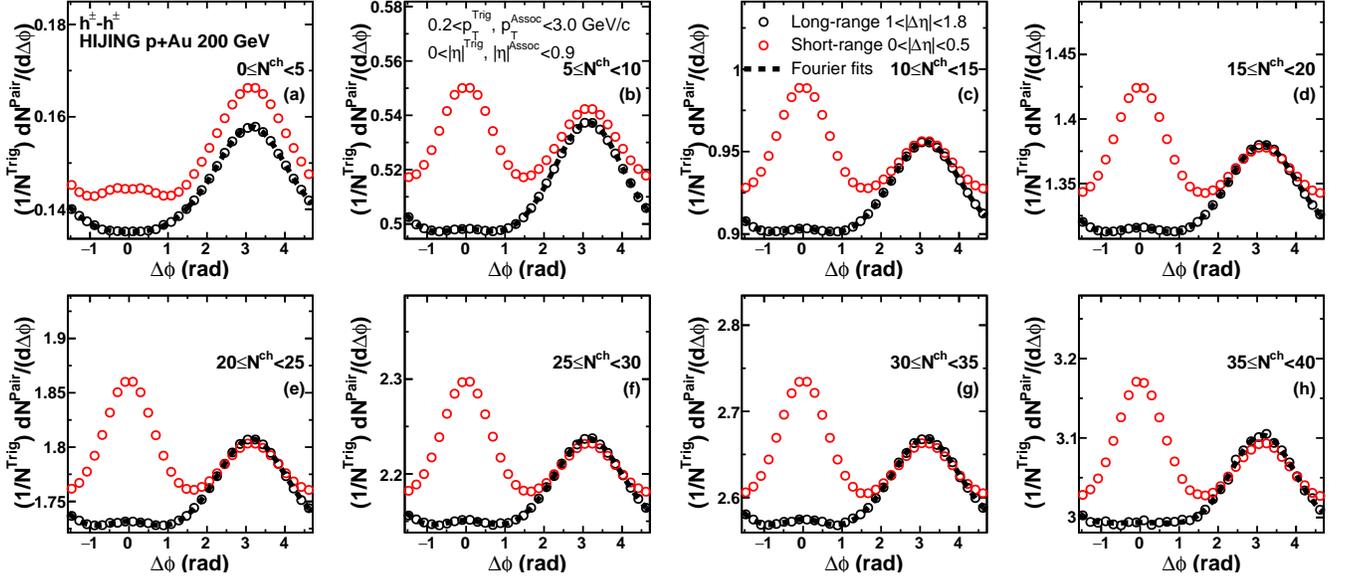


FIG. 23. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) ranges in p +Au collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING. Charged hadrons in $0.2 < p_T < 3$ GeV/ c and $|\eta| < 0.9$ are used for the correlation function. Each panel show a different multiplicity range, and the multiplicity is defined as the number of charged particles in $p_T > 0.2$ GeV/ c and $|\eta| < 0.9$.

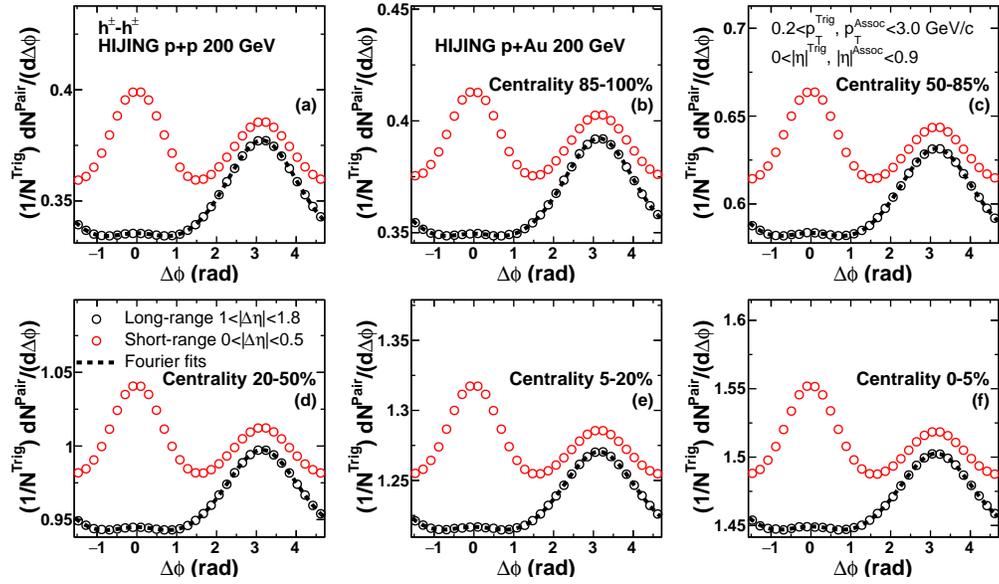


FIG. 24. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) ranges in $p+p$ (a) and $p+Au$ (b)–(f) collisions at $\sqrt{s_{NN}} = 200$ GeV from HIJING. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 0.9$ are used for the correlation function. Each panel of $p+Au$ collisions show a different centrality range, and the centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction).

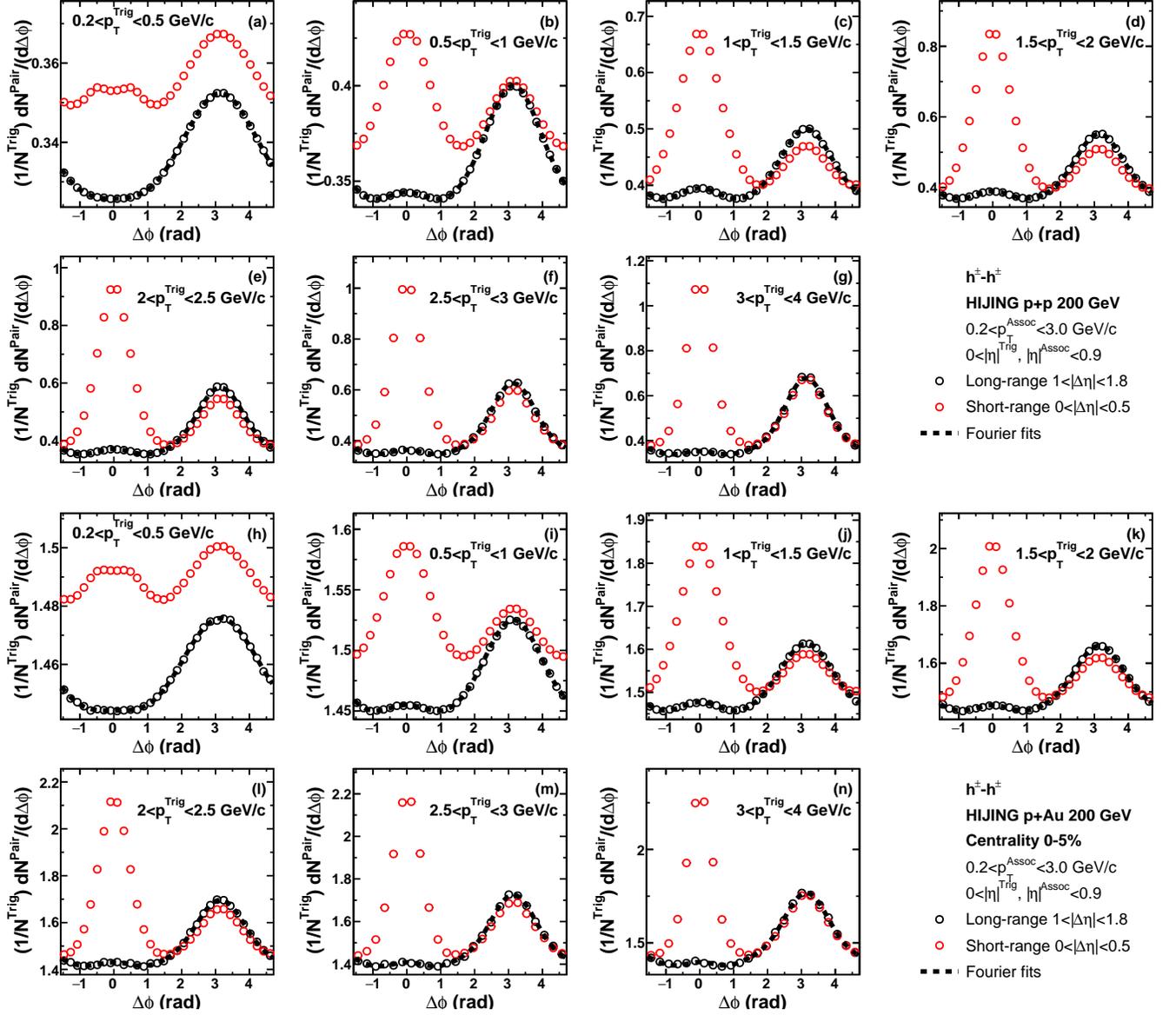


FIG. 25. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) ranges in $p+p$ (a)–(g) and 0–5% central $p+Au$ (h)–(n) collisions at $\sqrt{s_{NN}} = 200 \text{ GeV}$ from HIJING. Charged hadrons in $|\eta| < 0.9$ are used for the correlation function, and the p_T range of associated particle is $0.2 < p_T^{\text{Assoc}} < 3 \text{ GeV}/c$. Each panel of $p+Au$ collisions show a different p_T range of trigger particles.

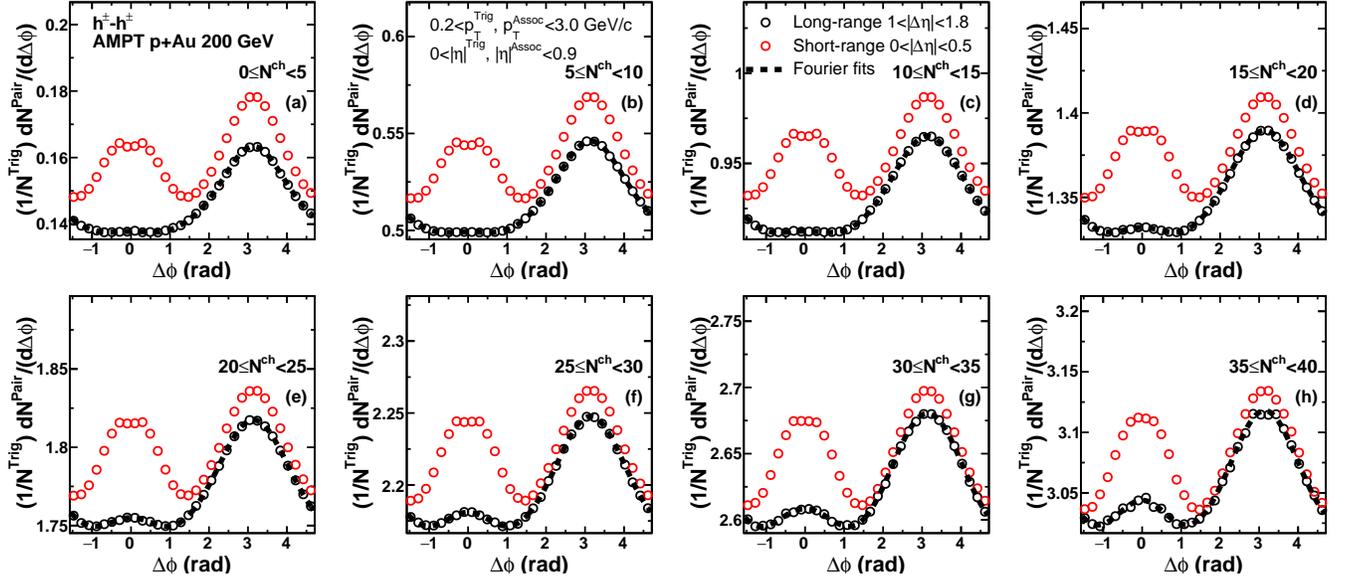


FIG. 26. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 0.5$) and long ($1.0 < |\Delta\eta| < 1.8$) ranges in p +Au collisions at $\sqrt{s} = 200$ GeV from AMPT. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 0.9$ are used for the correlation function. Each panel show a different multiplicity range, and the multiplicity is defined as the number of charged particles in $p_T > 0.2$ GeV/c and $|\eta| < 0.9$.

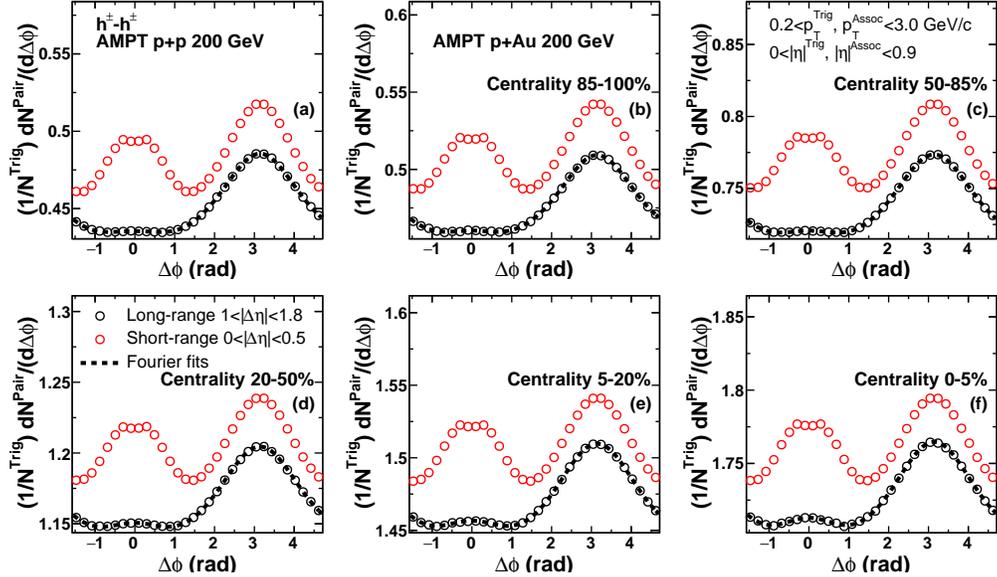


FIG. 27. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) ranges in p + p (a) and p +Au (b)–(f) collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT. Charged hadrons in $0.2 < p_T < 3$ GeV/c and $|\eta| < 0.9$ are used for the correlation function. Each panel of p +Au collisions show a different centrality range, and the centrality is defined as the number of charged particles in $-5.0 < \eta < -3.3$ (Au-going direction).

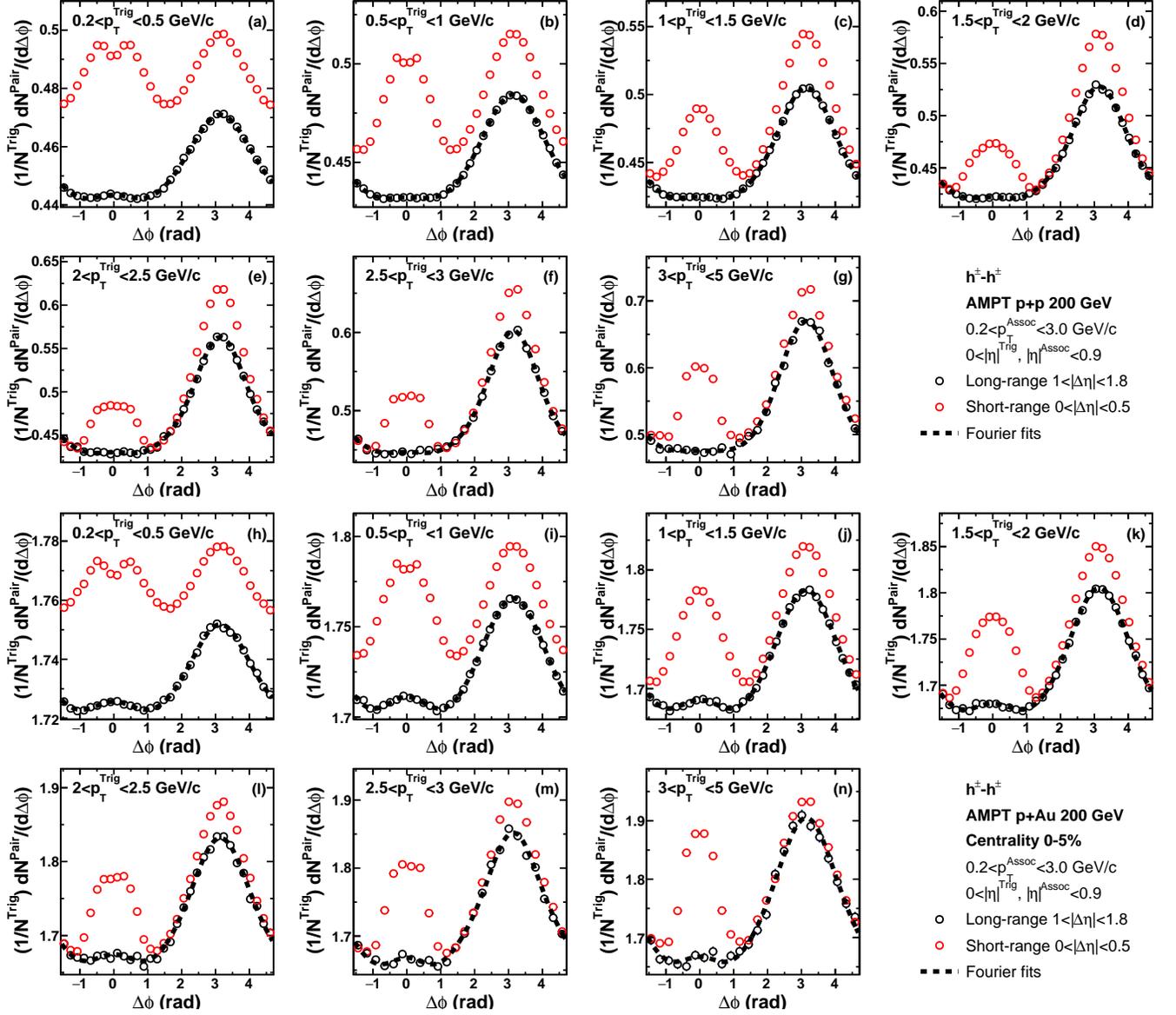


FIG. 28. Two particle $\Delta\phi$ correlation function at short ($|\Delta\eta| < 0.5$) and long ($1 < |\Delta\eta| < 1.8$) ranges in $p+p$ (a)–(g) and 0–5% central $p+Au$ (h)–(n) collisions at $\sqrt{s_{NN}} = 200$ GeV from AMPT. Charged hadrons in $|\eta| < 0.9$ are used for the correlation function, and the p_T range of associated particle is $0.2 < p_T^{\text{Assoc}} < 3$ GeV/c. Each panel of $p+Au$ collisions show a different p_T range of trigger particles.

-
- [1] Paul Romatschke and Ulrike Romatschke, “Relativistic Fluid Dynamics Out of Equilibrium – Ten Years of Progress in Theory and Numerical Simulations of Nuclear Collisions,” (2017), [arXiv:1712.05815 \[nucl-th\]](#).
- [2] Ulrich Heinz and Raimond Snellings, “Collective flow and viscosity in relativistic heavy-ion collisions,” *Ann. Rev. Nucl. Part. Sci.* **63**, 123–151 (2013), [arXiv:1301.2826 \[nucl-th\]](#).
- [3] Vardan Khachatryan *et al.* (CMS), “Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC,” *JHEP* **09**, 091 (2010), [arXiv:1009.4122 \[hep-ex\]](#).
- [4] James L. Nagle and William A. Zajc, “Small System Collectivity in Relativistic Hadronic and Nuclear Collisions,” *Ann. Rev. Nucl. Part. Sci.* **68**, 211–235 (2018), [arXiv:1801.03477 \[nucl-ex\]](#).
- [5] Ryan D. Weller and Paul Romatschke, “One fluid to rule them all: viscous hydrodynamic description of event-by-event central $p+p$, $p+Pb$ and $Pb+Pb$ collisions at $\sqrt{s} = 5.02$ TeV,” *Phys. Lett.* **B774**, 351–356 (2017), [arXiv:1701.07145 \[nucl-th\]](#).
- [6] C. Aidala *et al.* (PHENIX), “Creation of quark-gluon plasma droplets with three distinct geometries,” *Nature Phys.* **15**, 214–220 (2019), [arXiv:1805.02973 \[nucl-ex\]](#).
- [7] Michael Annan Lisa, Scott Pratt, Ron Soltz, and Urs Wiedemann, “Femtoscopy in relativistic heavy ion collisions,” *Ann. Rev. Nucl. Part. Sci.* **55**, 357–402 (2005), [arXiv:nucl-ex/0505014 \[nucl-ex\]](#).
- [8] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework,” *New computing techniques in physics research V. Proceedings, 5th International Workshop, AIHENP '96, Lausanne, Switzerland, September 2-6, 1996*, *Nucl. Instrum. Meth.* **A389**, 81–86 (1997).
- [9] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.* **178**, 852–867 (2008), [arXiv:0710.3820 \[hep-ph\]](#).
- [10] S. Voloshin and Y. Zhang, “Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions,” *Z. Phys.* **C70**, 665–672 (1996), [arXiv:hep-ph/9407282 \[hep-ph\]](#).
- [11] Miklos Gyulassy and Xin-Nian Wang, “HIJING 1.0: A Monte Carlo program for parton and particle production in high-energy hadronic and nuclear collisions,” *Comput. Phys. Commun.* **83**, 307 (1994), [arXiv:nucl-th/9502021 \[nucl-th\]](#).
- [12] Zi-Wei Lin, Che Ming Ko, Bao-An Li, Bin Zhang, and Subrata Pal, “A Multi-phase transport model for relativistic heavy ion collisions,” *Phys. Rev.* **C72**, 064901 (2005), [arXiv:nucl-th/0411110 \[nucl-th\]](#).
- [13] Georges Aad *et al.* (ATLAS), “Observation of Long-Range Elliptic Azimuthal Anisotropies in $\sqrt{s} = 13$ and 2.76 TeV pp Collisions with the ATLAS Detector,” *Phys. Rev. Lett.* **116**, 172301 (2016), [arXiv:1509.04776 \[hep-ex\]](#).
- [14] Morad Aaboud *et al.* (ATLAS), “Measurements of long-range azimuthal anisotropies and associated Fourier coefficients for pp collisions at $\sqrt{s} = 5.02$ and 13 TeV and $p+Pb$ collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector,” *Phys. Rev.* **C96**, 024908 (2017), [arXiv:1609.06213 \[nucl-ex\]](#).
- [15] Morad Aaboud *et al.* (ATLAS), “Correlated long-range mixed-harmonic fluctuations measured in pp , $p+Pb$ and low-multiplicity $Pb+Pb$ collisions with the ATLAS detector,” *Phys. Lett.* **B789**, 444–471 (2019), [arXiv:1807.02012 \[nucl-ex\]](#).
- [16] Vardan Khachatryan *et al.* (CMS), “Evidence for collectivity in pp collisions at the LHC,” *Phys. Lett.* **B765**, 193–220 (2017), [arXiv:1606.06198 \[nucl-ex\]](#).
- [17] Georges Aad *et al.* (ATLAS), “Observation of Associated Near-Side and Away-Side Long-Range Correlations in $\sqrt{s_{NN}} = 5.02$ TeV Proton-Lead Collisions with the ATLAS Detector,” *Phys. Rev. Lett.* **110**, 182302 (2013), [arXiv:1212.5198 \[hep-ex\]](#).
- [18] Betty Abelev *et al.* (ALICE), “Long-range angular correlations on the near and away side in p -Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV,” *Phys. Lett.* **B719**, 29–41 (2013), [arXiv:1212.2001 \[nucl-ex\]](#).
- [19] Shengli Huang (STAR), “Long-range collectivity in small collision systems with two- and four-particle correlations at STAR,” *Proceedings, 27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2018): Venice, Italy, May 14-19, 2018*, *Nucl. Phys.* **A982**, 475–478 (2019).
- [20] C. Aidala *et al.* (PHENIX), “Measurements of azimuthal anisotropy and charged-particle multiplicity in $d+Au$ collisions at $\sqrt{s_{NN}} = 200, 62.4, 39,$ and 19.6 GeV,” *Phys. Rev.* **C96**, 064905 (2017), [arXiv:1708.06983 \[nucl-ex\]](#).
- [21] J. L. Nagle and J. Orjuela Koop, “A Quasiparticle Transport Explanation for Collectivity in the Smallest of Collision Systems,” *Proceedings, 27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2018): Venice, Italy, May 14-19, 2018*, *Nucl. Phys.* **A982**, 455–458 (2019), [arXiv:1807.04619 \[nucl-th\]](#).
- [22] Zi-Wei Lin, “Evolution of transverse flow and effective temperatures in the parton phase from a multi-phase transport model,” *Phys. Rev.* **C90**, 014904 (2014), [arXiv:1403.6321 \[nucl-th\]](#).