LU-TP 20-34
June 2020

Principal Component Analyses of Final States in Relativistic Nuclear Collisions using PYTHIA/Angantyr

Nicoline Krogh Hemme

Department of Astronomy and Theoretical Physics, Lund University

Bachelor thesis supervised by Leif Lönnblad and Christopher Plumberg

LUND UNIVERSITY

Abstract

We develop a program to perform a principal component analysis of flow fluctuations using simulations of nuclear collisions produced by the PYTHIA/Angantyr event generator. We consider both p+p and p+Pb collision events and study the leading and subleading modes of the second flow harmonic. The analysis is carried out using two different definitions of the flow matrix which differ in their sensitivity to multiplicity fluctuations. We also study the effect on the flow fluctuations of two collective mechanisms modeled by Angantyr: string shoving and colour reconnection.

Popular Science Abstract

At the two particle accelerators located in New York and Switzerland, the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC), respectively, experiments are conducted in which atomic nuclei are collided at ultra-relativistic velocities. The aim of these experiments is to study nuclear matter under extreme conditions, i.e. extreme high temperatures and energy densities, similar to the conditions of our Universe shortly after its birth.

A particularly interesting result of these experiments is the observation of a collective, fluid-like motion known as *flow* [13][16][29]. The two colliding nuclei can each be thought of as a thin disk, and they can collide in such a way that the overlapping area forms an almond-like shape rather than a full circle. A consequence of this geometry is that the density gradients are larger along the short axis of the "almond" than the long axis. This means that particles emitted along the short axis are "pushed" with greater force than particles travelling along the almond's long axis, leading to more momentum being produced in the former direction than in the latter. This phenomenon can be observed using particle detectors and is known as anisotropic flow.

Flow is conventionally interpreted as the signature of a *quark-gluon plasma* (QGP), an extremely hot and dense state of matter, often envisioned as a "soup" of quarks and gluons, which are some of the most elementary particles we know of. According to the Big Bang Theory, our Universe existed in this state during the first few microseconds. It is therefore an exceptionally interesting state of matter to study. However, it still remains to be properly explored whether observables such as flow can be explained without a QGP.

Simulations of heavy-ion collisions provide a tool to probe the properties of the system by comparing the results to experimental data. The programs that simulate these events are known as event generators. The PYTHIA event generator [28], developed at Lund University, is based on a microscopic model of particle interactions and does not assume that the system reaches a state of QGP. This makes it a valuable tool for testing whether we have truly created a QGP in heavy-ion collisions.

In this project, a tool will be developed to provide a detailed flow-analysis of the final-state particles produced in PYTHIA simulations of relativistic nuclear collisions. The analysis will enable detailed comparisons of these simulations with corresponding experimental measurements. The purpose of this project is to expand the tool-box for future research to be done in this field.

Contents

1 Introduction

In the Standard Model of particle physics, the strong force is the force responsible for binding colour multiplets, such as quarks, together into colour singlets, e.g. hadrons, and is mediated by the colour octet, the gluon. The strength of the strong force is quantified by the Quantum Chromodynamics (QCD) coupling constant, α_{QCD}, and a particularly interesting characteristic of this coupling constant is that its magnitude continuously decreases with increasing momentum transfer, a phenomenon known as *asymptotic freedom*[18]. A consequence of this is that at sufficiently high energy densities, the strong interaction between particles becomes weak enough that a system's color degrees of freedom may become *deconfined*, meaning that they are no longer bound together into color singlets but are able to travel freely[25]. For this reason, nuclear matter existing in such a deconfined phase is known as *quark-gluon plasma* (QGP) and corresponds to the state of our Universe during the first few microseconds following the Big Bang.

The aim of heavy ion collision experiments is to study strongly-interacting matter under extreme conditions, with the possibility of recreating a thermalised, deconfined system, i.e. a QGP. Experimental results have revealed convincing evidence that the systems formed in these collisions undergo a collective, fluid-like expansion [13][16][29], a well-known characteristic of a thermalised system[21]. Moreover, results from hydrodynamical simulations have been able to describe this data quite well [13][23]. However, a relevant question still remains: can we have this collective behavior (or "collectivity") without a deconfined, thermalised medium, and can we explain it without appealing to the notion of a fluid like the QGP?

One of the most widely used tools for probing the properties of the matter created in nuclear collisions is known as *anisotropic flow* [23]. It is an experimental observable which measures the degree of azimuthal asymmetry in the final state momentum distribution of emitted particles. It is a particularly useful probe in that it is sensitive both to the system's fluctuating initial state and to the system's subsequent evolution. In Ref. [2] the idea to apply a Principal Component Analysis (PCA) to two-particle anisotropic flow observables was proposed as a way of isolating linearly independent fluctuation modes contained within the anisotropic flow signal. The PCA may thus be used to place powerful constraints on both the initial stages and subsequent evolution of high-energy nuclear collisions.

In this project, the code will be developed which can be interfaced with the PYTHIA/Angantyr event generator [28] to perform a PCA of two-particle azimuthal correlations. This will enable the study of flow fluctuations within PYTHIA/Angantyr. Angantyr is a recently developed extension to PYTHIA which focuses on improved modelling of heavy-ion collisions [6]. The framework of PYTHIA/Angantyr is based on perturbative QCD (pQCD) and string-string interactions and, in contrast to the hydrodynamic models mentioned above, does not assume the formation of a thermalised medium.

The goal of this project is therefore to provide a tool for probing the evolution of simulated high-energy nuclear collisions in the PYTHIA/Angantyr approach, as a way of ascertaining whether QGP formation is essential to our understanding of high-energy nuclear collisions.

The outline of this thesis is as follows. In Sec. 2 the basic ideas of anisotropic flow in relativistic nuclear collisions are introduced and the necessary formalism for building the PCA is presented. Sec. 3 introduces the toy model we use for simulating anisotropic flow and lists the details of the PYTHIA/Angantyr data selected for our PCA. The chosen method for error estimation is also briefly outlined. In Sec. 4, we present and discuss the results of our toy model calculations, along with those obtained from our simulations using PYTHIA/Angantyr. Finally, we draw some general conclusions in Sec. 5.

2 Formalism

2.1 Nuclear Collision Kinematics and Geometry

A nuclear collision at today's attainable energies is a relativistic system. Relativistic kinematics is therefore needed for describing the system. The first quantity to consider is the four-momentum $\mathbf{p}^\mu$. For the system where the beam is aligned with the z-axis, the four-momentum in natural units ($\hbar = c = 1$) is defined as

$$\mathbf{p}^\mu = (E, p_x, p_y, p_z) = (E, p_T \cos\phi, p_T \sin\phi, p_z) \tag{2.1}$$

where E is the energy of the particle, p_T is the transverse momentum defined as $p_T = \sqrt{p_x^2 + p_y^2}$, ϕ is the azimuthal angle and p_z is the longitudinal momentum. This geometry is depicted in Figure 1.

Another angle of interest is the reaction plane angle, ψ_{RP}. The reaction plane is spanned by the beam axis and the impact parameter, $\vec{b}$, which is defined as the vector connecting the centres of the colliding nuclei. ψ_{RP} denotes the angle between the reaction plane and x_{lab}.

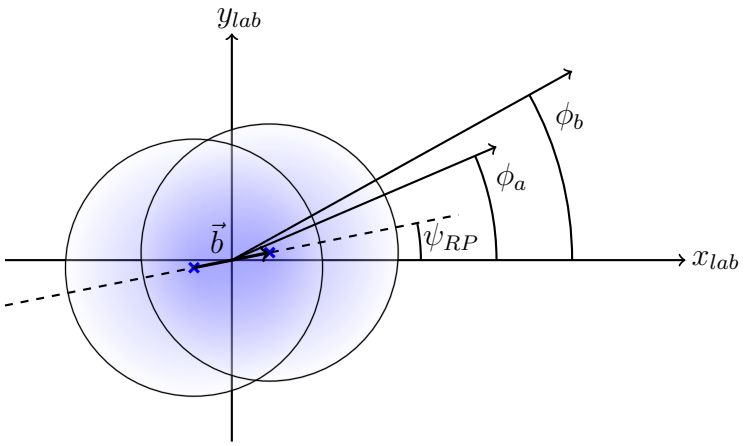

Figure 1: An illustration of a collision event, as seen when looking down the beam (z-axis). The circles signify two overlapping ions, with the blue crosses indicaitng the centers of the ions. The two vectors signify two outgoing particles. The dashed line is the reaction plane and the $\vec{b}$-vector is the impact parameter.

The impact parameter varies from event to event. The impact parameter can-

not be measured experimentally, but model studies show that it is correlated with experimental observables, such as the particle multiplicity and transverse energy [9].

Another useful parameter for a relativistic system is the rapidity, y, which can be defined as

$$y = \frac{1}{2}\ln\Big[\frac{E+p_z}{E-p_z}\Big] \tag{2.2}$$

The rapidity corresponds to the longitudinal velocity in a non-relativistic system. This can be seen by considering the limit of $v_z \ll 1$. Using the relation $p = Ev$ in natural units, the non-relativistic limit of Eq. 2.2 becomes:

$$y = \frac{1}{2}\ln\Big[\frac{E+Ev_z}{E-Ev_z}\Big] = \frac{1}{2}\ln\Big[\frac{1+v_z}{1-v_z}\Big] = \tanh^{-1}(v_z) \simeq v_z \text{ for } |v_z| \ll 1 \tag{2.3}$$

The benefit of rapidity over longitudinal velocity is that rapidity is an additive quantity under Lorentz boosts, in the same sense that angles are additive under rotations [9]. For a relativistic system, using the $p = Ev$ relation again along with $p_z = p\cos(\theta) \Rightarrow p_z = Ev\cos(\theta)$, where θ is the angular deviation of an outgoing particle from the z-axis, Eq. 2.2 can be rewritten to

$$y = \frac{1}{2}\ln\left[\frac{E+Ev\cos(\theta)}{E-Ev\cos(\theta)}\right] \simeq \frac{1}{2}\ln\Big[\frac{E+E\cos(\theta)}{E-E\cos(\theta)}\Big] = \frac{1}{2}\ln\Big[\frac{1+\cos(\theta)}{1-\cos(\theta)}\Big] \equiv \eta \text{ for } v \approx 1 \tag{2.4}$$

This is called the pseudorapidity. Note that it only depends on the angle of emission with respect to the beam axis and not on the mass or momentum of the particle. This quantity is particularly useful in actual experiments.

2.2 Anisotropic Flow

Consider a non-central collision between two nuclei, as illustrated in Figure 1 for the case of a symmetric collision between two heavy ions. In this case, the overlapping area between the two colliding nuclei will be of a roughly elliptical shape. This means that the collision system in the overlapping region will exhibit larger energy density gradients in directions (anti-)parallel to the impact parameter ("in-plane") than in those perpendicular to it ("out-of-plane"). As a result, the system will experience stronger collective expansion in-plane than out-of-plane, thereby generating an anisotropy ($\sim \cos(2\phi)$ oscillation) in the momentum distribution of particles produced by the collision.

Thus, geometric anisotropies in the system's initial state generate experimentally observable anisotropies in the momentum distributions of particles produced in nuclear collisions [2]. This phenomenon is known as *anisotropic flow*. For a given collision event, one may systematically identify these anisotropies in the single-particle (momentum-space) distribution by means of a Fourier expansion:

$$E_p\frac{dN}{d^3p} = \frac{dN}{p_T dp_T dy d\phi} = \frac{dN}{p_T dp_T dy}\sum_{n=-\infty}^{\infty} v_n e^{-in(\phi-\psi_n)} \tag{2.5}$$

The v_n coefficients are known as the *anisotropic flow coefficients* or *flow harmon-*

ics, and their magnitudes quantify the anisotropies present in the momentum-space distribution of a collision's particle production. The ψ_n phase is known as the *flow-plane angle* and denotes the orientation of the nth-order flow harmonic in the transverse plane. This thesis will focus on the harmonic with $n = 2$, which is known as the elliptic flow harmonic.

From Eq. (2.5) the Fourier coefficients can be found in terms of the single-particle distributions as functions of p_T and y:

$$v_n(p_T, y)e^{in\psi_n(p_T,y)} = \frac{\int_{-\pi}^{\pi} d\phi e^{i\phi} \frac{dN}{p_T dp_T dy d\phi}}{\int_{-\pi}^{\pi} d\phi \frac{dN}{p_T dp_T dy d\phi}} \equiv \left\{e^{i\phi}\right\}_{p_T,y} \tag{2.6}$$

For this thesis, the following notation will be used:

$$V_n(p_T, y) \equiv v_n(p_T, y)e^{in\psi_n(p_T,y)} \equiv \left\{e^{i\phi}\right\}_{p_T,y} \tag{2.7}$$

Here the curly brackets denote the average over a single event and a shorthand notation has been introduced in the last step, which emphasizes that the average over the azimuthal distribution is differential in p_T and y. Eq .(2.6) is the definition of *differential flow* [22]. A fully integrated definition of flow, the *integrated flow*, can similarly be written

$$V_n = \frac{\int_{-\pi}^{\pi} d\phi e^{i\phi} \frac{dN}{d\phi}}{\int_{-\pi}^{\pi} d\phi \frac{dN}{d\phi}} \equiv \frac{\int dy\, dp_T\, p_T \int_{-\pi}^{\pi} d\phi e^{i\phi} \frac{dN}{p_T dp_T dy d\phi}}{\int dy\, dp_T\, p_T \int_{-\pi}^{\pi} d\phi \frac{dN}{p_T dp_T dy d\phi}} \tag{2.8}$$

$$\equiv \left\{e^{i\phi}\right\} \tag{2.9}$$

Unfortunately, the single-particle distribution is dependent on the reaction plane angle, whose determination is sensitive to finite-statistical fluctuations in the number of particles emitted on an event-by-event basis [27]. To avoid such uncertainties in the reaction plane determination, the distribution of pairs of particles within the same event can be considered instead. The Fourier expansion of the pair-particle distribution is defined in analogy with Eq. (2.5):

$$\frac{dN_{\text{pairs}}}{d\mathbf{p}_a d\mathbf{p}_b} = \sum_{n=-\infty}^{\infty} V_{n\Delta}(p_a, p_b)e^{-in(\phi_a-\phi_b)} \tag{2.10}$$

where $\mathbf{p} \equiv (p_T, y, \phi)$ and $p \equiv (p_T, y)$. The Fourier coefficients for the pair-particle distribution of a single event, i, are given by:

$$V_{n\Delta}^{(i)}(\mathbf{p}_a, \mathbf{p}_b) = \left\{e^{-in(\phi_a-\phi_b)}\right\}_{p_a,p_b}^{(i)} \tag{2.11}$$

The flow matrix, constructed as an average over many events is thus given by

$$V_{n\Delta}(\mathbf{p}_a, \mathbf{p}_b) = \left\langle \left\{e^{-in(\phi_a-\phi_b)}\right\}_{p_a,p_b} \right\rangle \tag{2.12}$$

where

$$\langle X \rangle \equiv \frac{1}{N_{\text{events}}} \sum_{i=1}^{N_{\text{events}}} X_i \tag{2.13}$$

defines the "ensemble average" of X over the set of collision events being considered. The flow matrix (2.12) can be estimated by the following equation

$$V_{n\Delta}(\mathbf{p}_a, \mathbf{p}_b) = \left\langle \frac{\sum_{\text{pairs}} e^{-in(\phi_a - \phi_b)}}{N_{\text{pairs}}(\mathbf{p}_a, \mathbf{p}_b)} \right\rangle \tag{2.14}$$

However, due to finite statistics, particularly important in experiments, another definition of the flow matrix is often preferred by experimentalists [24][13], and is presented here

$$V^N_{n\Delta}(\mathbf{p}_a, \mathbf{p}_b) = \left\langle \sum_{\text{pairs}} e^{-in(\phi_a - \phi_b)} \right\rangle \tag{2.15}$$

This method requires compensation for the extra factors of particle numbers, which will be explained below in the *Principal Component Analysis* section. Both definitions of the flow matrix will be explored in this thesis.

2.3 Event-by-event Fluctuations

The initial conditions of a collision, such as the impact parameter and the transverse density profile of the system, fluctuate from event to event due in large part to the random distribution of nucleons within each nucleus at the moment of collision. Even events with identical impact parameter will have fluctuations in the initial density profile due to such quantum fluctuations [22]. These event-by-event fluctuations of the initial state induce corresponding fluctuations in the anisotropies of the final-state momentum spectra [24].

The Fourier coefficients in Eq. (2.11) can be expressed as the covariance matrix:

$$V_{n\Delta}(\mathbf{p}_a, \mathbf{p}_b) = \langle V_n^*(\mathbf{p}_a) V_n(\mathbf{p}_b) \rangle \tag{2.16}$$

In the absence of event-by-event fluctuations, the covariance matrix factorises as

$$V_{n\Delta}(\mathbf{p}_a, \mathbf{p}_b) = \sqrt{\langle |V_n(\mathbf{p}_a)|^2 \rangle} \sqrt{\langle |V_n(\mathbf{p}_b)|^2 \rangle} \tag{2.17}$$

and the covariance matrix will only have one non-vanishing eigenvalue [24], which corresponds to the anisotropic flow discussed above. However, in the presence of event-by-event fluctuations this factorisation breaks and the covariance matrix will have more than one non-vanishing eigenvalue, with each eigenvalue corresponding to a linearly independent fluctuation mode [2].

A principal component analysis of the event-by-event fluctuations can isolate these fluctuation modes, organised by size, and thereby reveal the most important contributions to the covariance matrix [2].

These fluctuation modes are sensitive in non-trivial ways to both the initial conditions of nuclear collisions as well as the nature of their subsequent evolution.

For this reason, they may also provide valuable into the properties of both the initial conditions and the mechanisms which drive the evolution.

2.4 Principal Component Analysis

The covariance matrix Eq. (2.16) is symmetric and approximately real after averaging over many events [22]. By the spectral theorem, therefore, it may be decomposed into positive semi-definite eigenvalues and real eigenvectors in the following way:

$$V_{n\Delta}(\mathbf{p}_a, \mathbf{p}_b) = \sum_{\alpha=1}^{\infty} \lambda_n^{(\alpha)} \psi_n^{(\alpha)}(\mathbf{p}_a) \psi_n^{(\alpha)}(\mathbf{p}_b) \tag{2.18}$$

$$= \sum_{\alpha=1}^{\infty} V_n^{(\alpha)}(\mathbf{p}_a) V_n^{(\alpha)}(\mathbf{p}_b) \tag{2.19}$$

where $\lambda_n^{(\alpha)}$ are the eigenvalues of $V_{n\Delta}$, in descending order, $\lambda_n^{(\alpha)} > \lambda_n^{(\alpha+1)}$, and $\psi_n^{(\alpha)}(\mathbf{p})$ are the normalized eigenvectors. The principal components are then, in accordance with Eq. (2.14), defined to be

$$V_n^{(\alpha)}(\mathbf{p}) = \sqrt{\lambda_n^{(\alpha)}} \psi_n^{(\alpha)}(\mathbf{p}) \tag{2.20}$$

However, when considering the other method of Eq. (2.15), an extra factor of $\langle N(\mathbf{p}) \rangle^{-1}$ is required in order to properly normalize the extracted flow modes. The principal components are then defined by

$$V_n^{N(\alpha)}(\mathbf{p}) = \sqrt{\lambda_n^{(\alpha)}} \psi_n^{(\alpha)}(\mathbf{p}) / \langle N(\mathbf{p}) \rangle \tag{2.21}$$

2.5 The PYTHIA/Angantyr Event Generator

PYTHIA 8 is a general purpose, Monte Carlo event generator designed to simulate the physics of particle collisions carried out at relativistic particle colliders such as RHIC and LHC. Within PYTHIA, the partonic interactions are modelled by perturbative QCD while the hadronization is based on the Lund String Model, which models high-energy QCD interactions by strings whose energy density is stored in the form of an effective tension. For nearly four decades, PYTHIA has provided successful descriptions of a wide range of both elementary particle and hadronic collisions, including e^+e^-, pp, and $p\bar{p}$.

More recently, PYTHIA has been extended to the description of particle-nucleus (pA) and nucleus-nucleus (AA) collisions by treating the multiple nucleon-nucleon interactions in such collisions in accordance with the same Lund String Model. The extended model has been dubbed the Angantyr model. One of the most important recent developments in this respect has been the application of PYTHIA/Angantyr to the description of collective phenomena [1] such as anisotropic flow [10], quarkonium production and suppression [15], and long-range rapidity correlations [19, 11, 14]. Several novel mechanisms have been implemented within PYTHIA/Angantyr in order to account for such collective phenomena. One of the main features in-

troduced is *rope hadronization*, in which strings overlapping in transverse space can interact to form a "rope". As part of the rope hadronization mechanism, other types of interactions have been introduced, two of which will be explored in this project, namely *string shoving* and *colour reconnection*.

These mechanisms, which will be briefly discussed below, may have quantifiable implications for experimental observables which probe collectivity. This includes, in particular, observables associated to anistropic flow, such as the flow matrix $V_{n\Delta}$ and its eigenmodes, which may be extracted by principal component analyses as described above. In this thesis, we will consider the implications of PYTHIA/Angantyr's collective mechanisms for PCAs performed on this flow matrix.

2.5.1 String Shoving

In the Lund String model, a colour field can be formed between a $q\bar{q}$-pair, which can fragment into two colour singlets (hadrons) when the energy stored in a string is sufficient for the creation of a new $q\bar{q}$-pair. These colour fields can be interpreted as QCD flux tubes and stretch out in both the longtitudinal and transverse direction. When a $q\bar{q}$-pair is produced, the flux tube connecting them is compressed in the transverse directions, and as the partons move apart, the tube is expanded. In dense environments, this leads to flux tubes overlapping and interacting by a repulsion in the transverse direction, referred to as *shoving* [5].

2.5.2 Colour Reconnection

The interacting strings of the rope can be of different colour configurations, and as they overlap they can interact by a rearrangement to other colour configurations, determined by the group properties of SU(3) [3]. This is the process referred to as *colour reconnection* (CR). Some of the configurations possible with CR are the *junction configurations*, where one or more junctions are found along the string. Each of these junctions allow for the creation of additional baryons [4].

3 Method

3.1 Acquiring Data

3.1.1 Toy Model

To illustrate the functionality of the PCA, a toy model with well defined expected behaviour has been constructed. This model simulates elliptic flow in fluctuating nuclear collisions in a way which can produce a non-factorisable flow matrix. The toy model mimics elliptic flow by randomly generating particles whose momentum-space azimuthal angle is distributed according to a (properly normalized) cosine distribution

$$\phi \sim 1 + 2v_2(p_T)\cos[2(\phi - \psi_2(p_T))] \tag{3.1}$$

with

$$v_2(p_T) = \frac{\bar{v}_2 p_T}{\bar{p}_T + p_T} \tag{3.2}$$

and

$$\psi_2(p_T) = \psi_{RP} + k p_T \tag{3.3}$$

Here, $\bar{v}_2$ and $\bar{p}_T$ are model parameters which set the effective elliptic flow magnitude and typical p_T scale, respectively. In this analysis, the values $\bar{v}_2 = 0.25$ and $\bar{p}_T = 1$ GeV have been used. Additionally, the fluctuating parameter k in Eq. (3.3) is included in order to induce a p_T dependence in the second-order flow plane angle ψ_2 which mimics the usual effects of event-by-event fluctuations in nuclear collisions. This p_T dependence is responsible for the breaking of flow factorization in these collisions, as will be shown in Sec. 4 by comparing the results for $k = 0$ with those for $k \neq 0$ [22]. In the latter case, k is sampled on an event-by-event basis from a uniform distribution on the interval [0, 0.1] in units of 1/GeV. ψ_{RP} of course also fluctuates uniformly from 0 to 2π on an event-by-event basis.

The p_T distribution is chosen to be exponential with scale parameter $\bar{p}_T = 1$ GeV, as above. Additionally, since the flow matrix is not constructed differentially in rapidity y or pseudorapidity η in this project, the longitudinal structure is less important than the p_T and ϕ distributions. Therefore, in the interest of simplicity, each particle's η is chosen according to a normal distribution with zero mean and a standard deviation of 2 units in pseudorapidity. This is sufficient to generate semi-realistic flow distributions.

3.1.2 PYTHIA/Angantyr

In this project two different types of collision systems are considered: p+p and p+Pb collisions at a beam energy of $\sqrt{s_{NN}} = 14$ TeV and $\sqrt{s_{NN}} = 5.02$ TeV, respectively. For each collision type, four different models are explored. These are as listed in Table 1 below. No retuning against experimental data has been performed for the models in this project.

Table 1: The four different models considered in this thesis.

	String Shoving	**Colour Reconnection**
Model 1	Off	Off
Model 2	Off	On
Model 3	On	Off
Model 4	On	On

Events are simulated in four different multiplicity classes on the basis of the charged particle multiplicity, N_{ch}. The four classes are as follows: $120 \leq N_{ch} < 150$, $150 \leq N_{ch} < 185$, $185 \leq N_{ch} < 220$ and $220 \leq N_{ch} < 260$; they correspond to the multiplicity classes used for p+Pb collisions in a recent PCA analysis performed by the CMS collaboration [13]. Note that while all charged particle species are used in determining the multiplicity class, only charged pions are used in constructing the flow matrix in this analysis.

For p+p collisions, 10^6 events are generated for each model and each multiplicity class. Similarly, for p+Pb collisions, 10^5 events are generated in each case.

3.2 The Flow Matrix and Eigendecomposition

The flow matrix of this project is differential in p_T and cover a p_T range of $0.3 < p_T < 3.0$ GeV/c, where the number of (uniformly spaced) p_T bins $n_{p_T} = 6$.

Since there us no binning done in η, the flow matrix is effectively integrated over this direction. However, a rapidity gap of $|\Delta\eta| > 2.0$ is imposed on the pairs used to construct the flow matrix. This is a common procedure to minimise non-flow contributions to the flow signal [12]. Such contributions can arise from jets and resonance decays, in which many particles emerge with similar rapidity and outgoing angle, and can therefore dominate the correlations at small $\Delta\eta$ and $\Delta\phi$.

After the flow matrix is constructed, we use the GSL library [20] to perform the eigendecomposition and the extraction of the flow modes.

3.3 Error Estimation

The errors of the flow matrix and eigenvalues and -vectors are estimated using the *Jackknife* method [26]. For a dataset of M values the method requires M unique Jackknife samples, each constructed by deleting a single value from the dataset. In this case, $M = N_{\text{events}}$. Define the full dataset, $\mathbf{x}$, and the ith unique sample, $\mathbf{x}_i$ by

$$\mathbf{x} = \{x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_M\} \tag{3.4}$$

$$\mathbf{x}_i = \{x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_M\} \tag{3.5}$$

Each element x_i corresponds to the evaluation of a fluctuating quantity (e.g., the subleading eigenvalue λ_1) in a single event. Each eigenvalue, eigenvector element, and p_T-bin multiplicity is taken to be an independently fluctuating quantity in this regard.

The variance squared, σ^2, is then defined as

$$\sigma^2 = \frac{M-1}{M} \sum_{i=0}^{M} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^2 \tag{3.6}$$

where $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{x}}$ represent the arithmetic averages of the elements in $\mathbf{x}_i$ and $\mathbf{x}$, respectively.

The uncertainties on the eigenvalues $\lambda_n^{(\alpha)}$ and the eigenmodes $\psi_n^{(\alpha)}$[1] are evaluated using the jackknife approach. The corresponding uncertainties are denoted by $\delta\lambda_n^{(\alpha)}$ and $\delta\psi_n^{(\alpha)}$.

[1]Note that the eigenmodes $\psi_n^{(\alpha)}$ and $\psi_n^{N(\alpha)}$ should be distinguished from the flow-plane angles ψ_n.

Using this notation, the estimated uncertainty of $V_n^{(\alpha)}(\mathbf{p})$ is

$$\delta V_n^{(\alpha)}(\mathbf{p}) = \left|\sqrt{\lambda_n^{(\alpha)}}\psi_n^{(\alpha)}\right| \sqrt{\left(\frac{\delta\lambda_n^{(\alpha)}}{2\lambda_n^{(\alpha)}}\right)^2 + \left(\frac{\delta\psi_n^{(\alpha)}}{\psi_n^{(\alpha)}}\right)^2} \tag{3.7}$$

where the momentum $\mathbf{p}$ is held fixed.

A similar result holds for the $V_n^{N(\alpha)}(\mathbf{p})$ of Eq. (2.21). In this case, the error on $\langle N(\mathbf{p})\rangle$ must also be estimated and is determined by the standard error

$$\delta\langle N(\mathbf{p})\rangle = \frac{\sigma_p}{\sqrt{N_{\text{events}}}} \tag{3.8}$$

where $\sigma_{N(\mathbf{p})}$ is the standard deviation

$$\sigma^2_{N(\mathbf{p})} = \frac{\sum(N(\mathbf{p})_i - \langle N(\mathbf{p})\rangle)^2}{N_{\text{events}} - 1} \tag{3.9}$$

This gives the final error estimation

$$\delta V_n^{N(\alpha)}(\mathbf{p}) = \left|\frac{\sqrt{\lambda_n^{N(\alpha)}}\psi_n^{N(\alpha)}}{\langle N\rangle}\right| \sqrt{\left(\frac{\delta\lambda_n^{N(\alpha)}}{2\lambda_n^{N(\alpha)}}\right)^2 + \left(\frac{\delta\psi_n^{N(\alpha)}}{\psi_n^{N(\alpha)}}\right)^2 + \left(\frac{-\delta\langle N\rangle}{\langle N\rangle}\right)^2} \tag{3.10}$$

and the $\mathbf{p}$ dependence is again held fixed and suppressed.

In both Eqs. (3.7) and (3.10), any possible covariance between the various uncertainties has been neglected in the interest of simplicity. This approximation could be relaxed in a future analysis, but is good enough to give a rough feeling for the importance of statistical uncertainties on the results that will be presented here.

4 Results

In Figure 2, the results of the toy model (for 10^5 events) are depicted. They confirm the role of p_T dependent flow fluctuations in the factorization breaking of the flow matrix. The leading mode for both definitions of the flow matrix increases towards high p_T. Regarding the subleading modes, it can be seen that the datasets from a distribution without p_T-dependent fluctuations (subplot (a)) generate an almost negligible subleading mode. On the other hand, the distribution with p_T-dependent fluctuations (subplot(b)) has a non-vanishing subleading mode, clearly indicating a breakdown of factorization. Finally, while the leading modes of $V_2^{(\alpha)}$ are always positive, the subleading modes of $V_2^{(\alpha)}$ are partially positive and negative, as must be the case in order for the eigenmodes to fullfill the expected orthonormality relation [24].

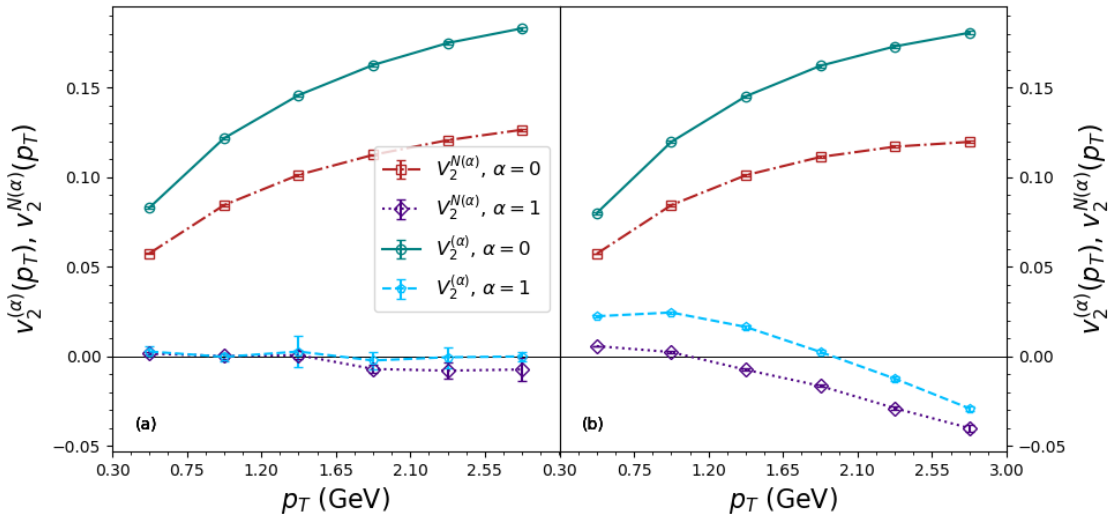

Figure 2: Leading ($\alpha = 0$) and subleading ($\alpha = 1$) modes from our toy model-generated data of 10^5 events. Subfigure (a) does not include a p_T dependence in the second-order flow plane angle ψ_2, while subfigure (b) does.

Figures 3-4 show the leading mode results of the p+p collisions. Similarly, Figures 5-6 show the leading mode results of the p+Pb collisions.

Immediately noticeable in the leading modes of the p+p, as well as the p+Pb collision events seen below, is the difference in behaviour between the two definitions of the flow matrix, with $V_2^{(\alpha)}$ decreasing at larger p_T-values while $V_2^{N(\alpha)}$ continues to increase. While $V_2^{N(\alpha)}$ has the advantage over $V_2^{(\alpha)}$ that it gives more weight to events with more pairs, thereby reducing statistical uncertainties, it is sensitive to the effects of multiplicity fluctuations [24]. $V_2^{(\alpha)}$ may still be sensitive to some multiplicity fluctuations, but it eliminates a trivial and significant background noise and represents a more direct estimation of the anisotropic flow fluctuations.

An interesting behaviour is exhibited by Model 2 and Model 4. They tend to have small values in the $V_2^{N(\alpha)}$ definition, but larger values in the $V_2^{(\alpha)}$ definition. Both of these models have CR turned on and these results could appeal to the idea that CR may be more capable of generating anisotropic flow fluctuations than string shoving alone or than no additional mechanisms.

Model 3 slightly exhibits an exact opposite behaviour, having the larger values in the $V_2^{N(\alpha)}$ definition and the smaller values in the $V_2^{(\alpha)}$ definition. This could be a result of a higher sensitivity to the multiplicity in the string shoving mechanism than the remaining mechanisms, or possibly an indication that the string shoving mechanism may generate more multiplicity fluctuations than the remaining models.

However, due to the lack of tuning in this project, the observations regarding the different behaviours of the different models (the observations mentioned in the previous two paragraphs) should not be taken as conclusive statements. Enabling/disabling different mechanisms in these simulations, such as colour reconnection, can affect other parameters and ultimately affect observables such as charged multiplicity in a way which is undesired and which can be fixed by performing a tuning [7][17].

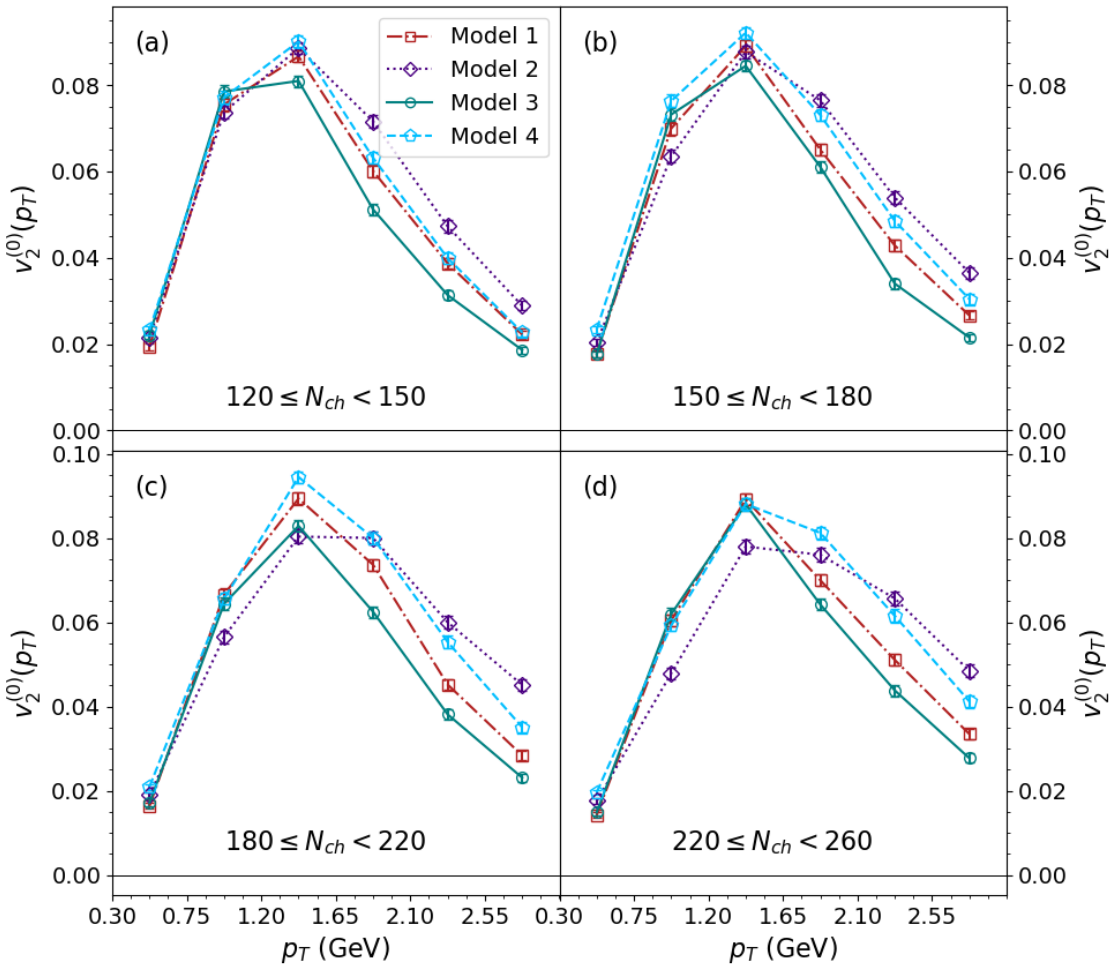

Figure 3: Leading modes for pp collisions as defined by Eq. (2.20). Error bars represent statistical uncertainties.

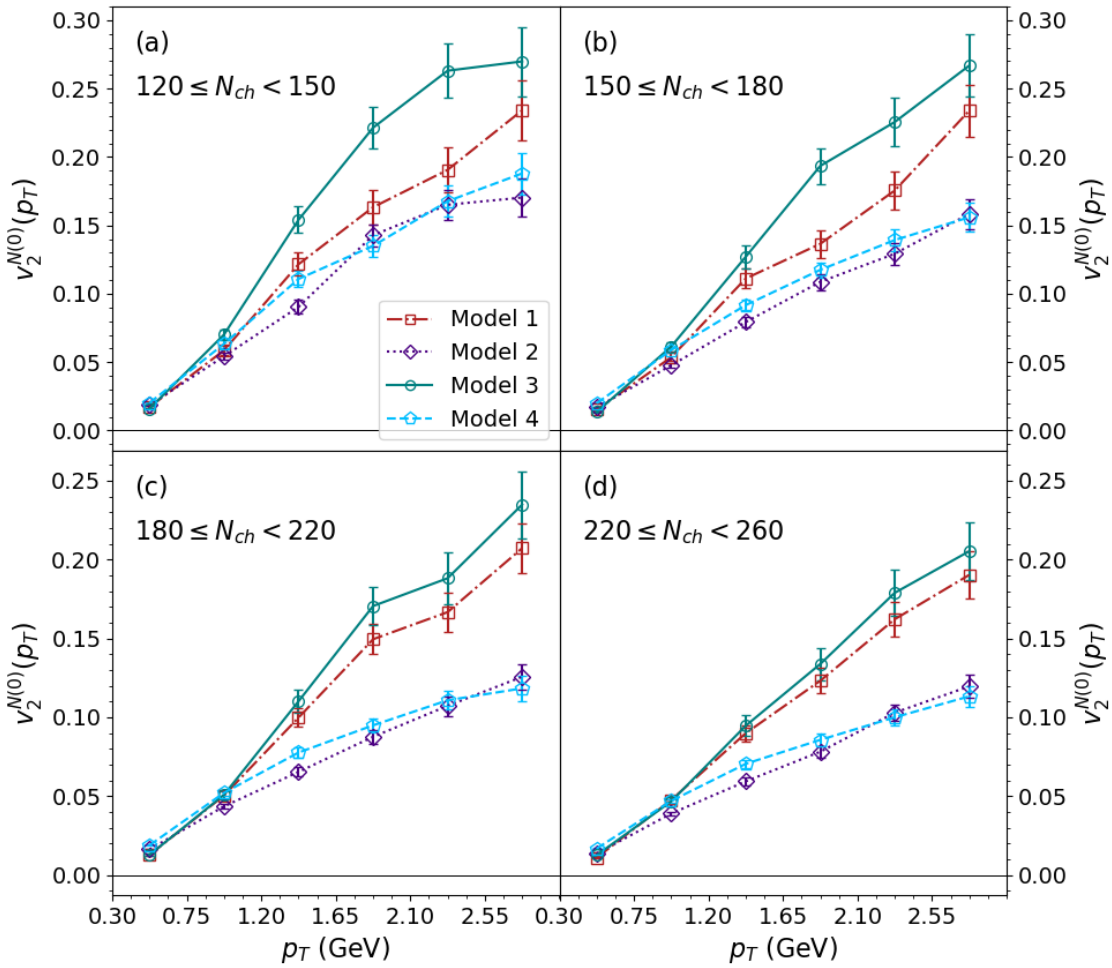

Figure 4: Leading modes for pp collisions as defined by Eq. (2.21). Error bars represent statistical uncertainties.

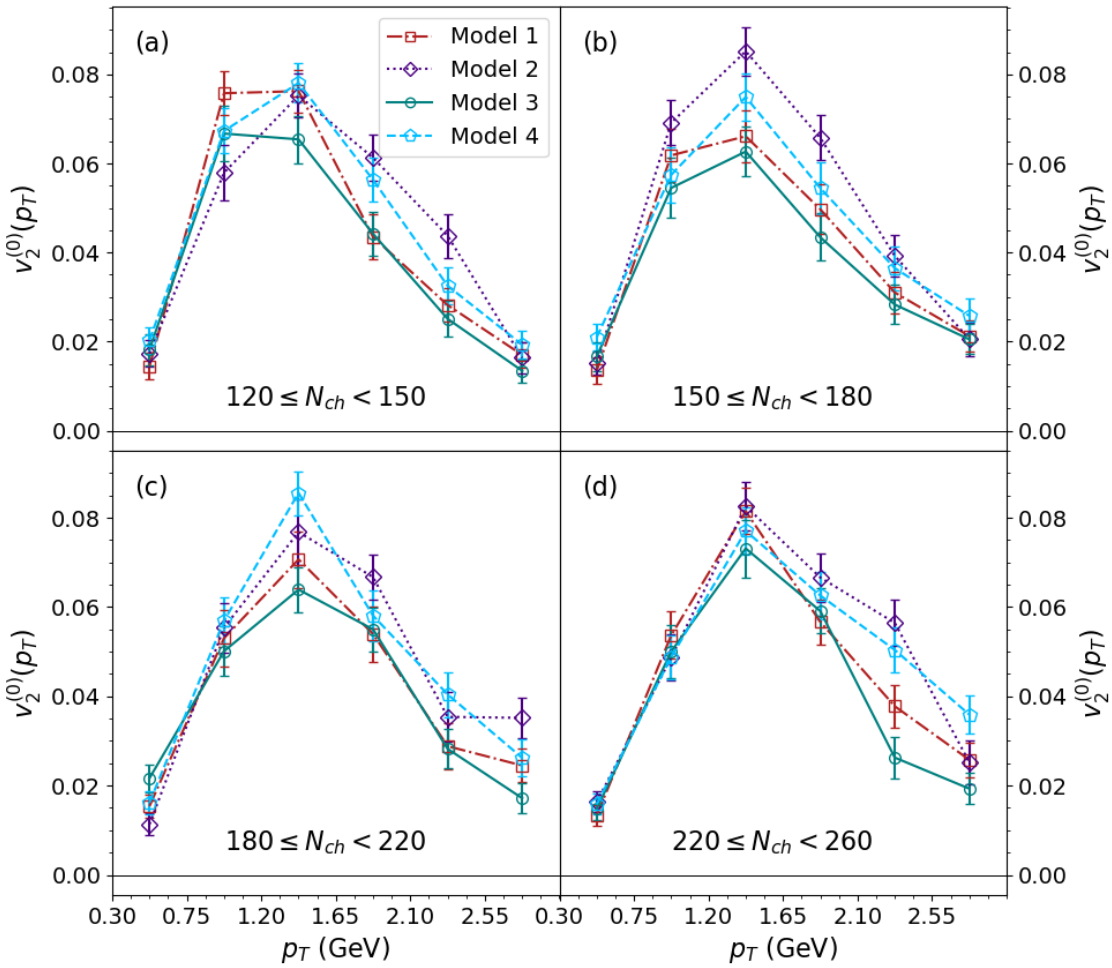

Figure 5: Leading modes for pPb collisions as defined by Eq. (2.20). Error bars represent statistical uncertainties.

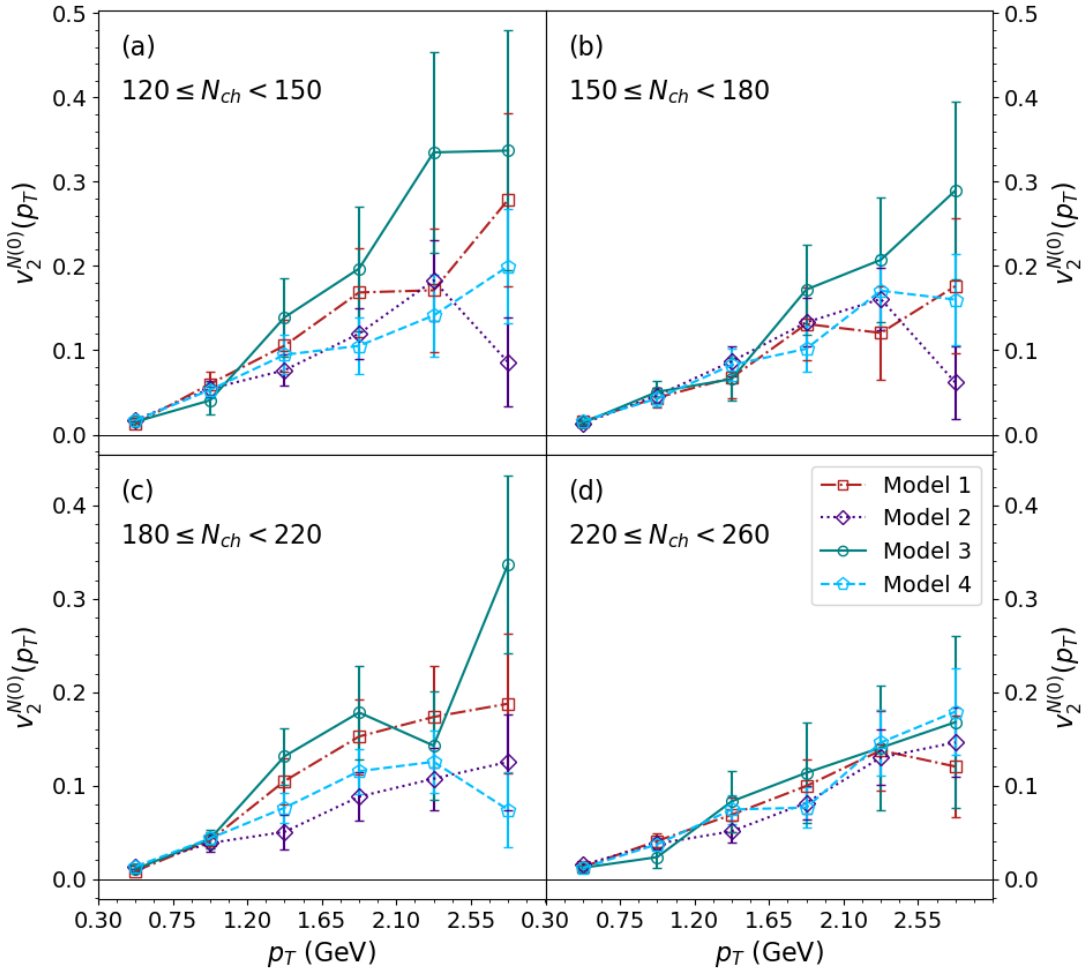

Figure 6: Leading modes for pPb collisions as defined by Eq. (2.21). Error bars represent statistical uncertainties.

The error bars of the leading modes of $V_2^{(\alpha)}$ as well as $V_2^{N(\alpha)}$ for the p+p collision events are virtually negligible. As expected, however, due to statistical fluctuations, the error bars of the subleading modes are significantly larger. Figures 7-10 show the subleading modes of the p+p and p+Pb collisions.

Regarding the orthogonality of the principal components, the $V_2^{(\alpha)}$ eigenmodes are orthogonal to an approximation, while the $V_2^{N(\alpha)}$ eigenmodes are not, clearly seen by their failure to fulfill the orthonormality relation.

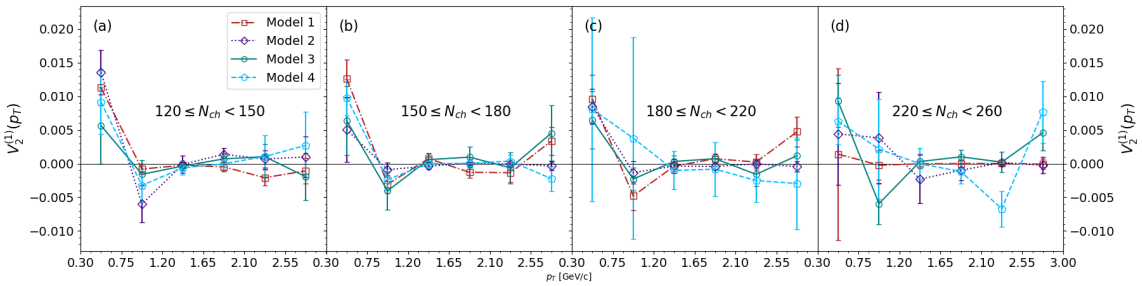

Figure 7: Subleading modes for pp collisions as defined by Eq. (2.20). Error bars represent statistical uncertainties.

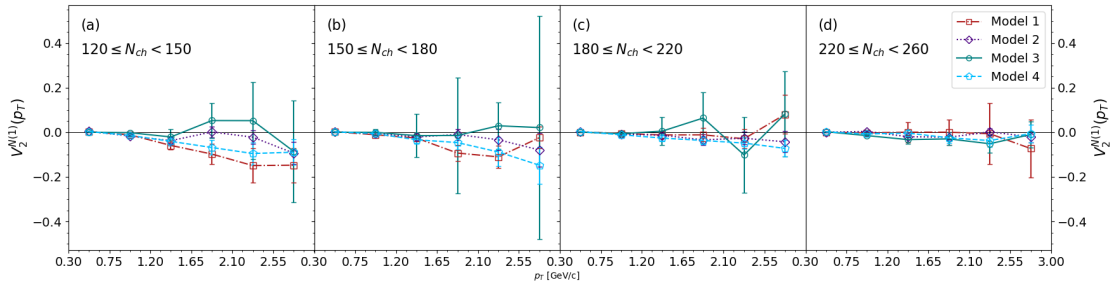

Figure 8: Subleading modes for pp collisions as defined by Eq. (2.21). Error bars represent statistical uncertainties.

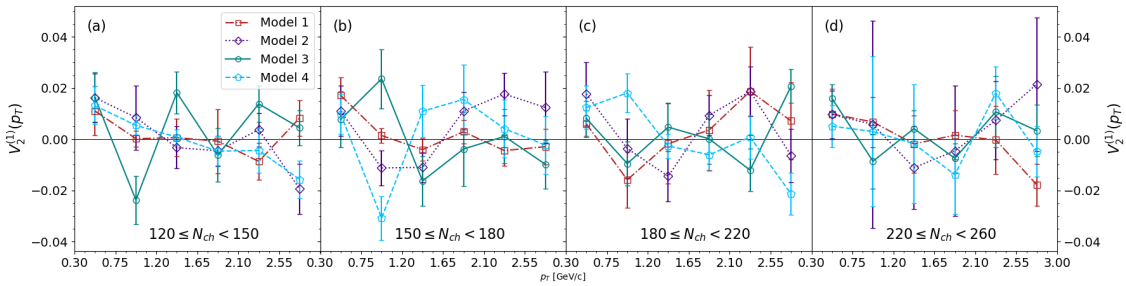

Figure 9: Subleading modes for pPb collisions as defined by Eq. (2.20). Error bars represent statistical uncertainties.

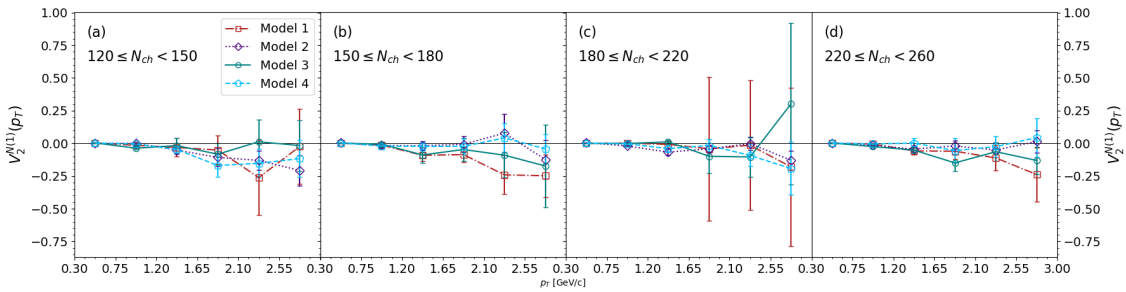

Figure 10: Subeading modes for pPb collisions as defined by Eq. (2.21). Error bars represent statistical uncertainties.

5 Conclusion

In this project we developed a code that applies a PCA to two-particle azimuthal corelations in nuclear collisions, with the option of using an alternative definition of the flow matrix often preferred by experimentalists. We tested our formalism on a toy model simulating elliptic flow. We also applied it to data generated by PYTHIA/Angantyr and studied the leading and subleading mode of the elliptic flow of generated p+p and p+Pb collision events. We found that the alternative flow matrix definition contains a trivial multiplicity fluctuation background. We showed that the alternative definition gives principal components that are not orthogonal.

The leading mode of $V_2^{N(\alpha)}$ of the p+Pb collision events exhibits similar trends to what has been calculated from CMS p+Pb collisions at the same multiplicity classes [13]. This correlation could be due to similar magnitudes of multiplicity fluctuations, rather than similar flow fluctuations. It would therefore be interesting to extend this research by performing a similar PCA (using the $V_2^{(\alpha)}$ definition) on the CMS data, which would remove the background for a better estimation of the actual flow fluctuation, to see whether and to what extent those results would agree with the calculations presented here.

Considering Pb+Pb collision events would be another interesting extension of this research. In particular, the geometry and centrality fluctuations of such collision events may play a significant role, as compared to the p+p and p+Pb collision events.

The PCA constructed in this project should also be able to perform a PCA on the triangular flow, v_3, although it has not been tested here. Extending this study to the triangular flow could also be exceptionally interesting as the event-by-event fluctuations have a particularly important effect on odd harmonics [10].

This thesis has shown the functionality of the constructed PCA and provided further grounds for the discussion of the relevance of multiplicity fluctuations PCAs using the $V_2^{N(\alpha)}$ flow matrix definition. There are many ways to extend on this research. The simulations studied here are hopefully only the beginning.

Acknowledgements

Computing resources from both the Minnesota Supercomputing Institute (MSI) at the University of Minnesota and the Ohio Supercomputer Center [8] are gratefully acknowledged.

References

[1] J. Adolfsson et al. "QCD Challenges from Pp to A-A Collisions". In: (Mar. 24, 2020). arXiv: 2003.10997.

[2] Rajeev S. Bhalerao et al. "Principal Component Analysis of Event-by-Event Fluctuations". In: (Oct. 28, 2014). DOI: 10.1103/PhysRevLett.114.152301.

[3] Christian Bierlich. "Hadronisation Models and Colour Reconnection". In: (June 30, 2016).

[4] Christian Bierlich and Jesper Roy Christiansen. "Effects of Color Reconnection on Hadron Flavor Observables". In: *Phys. Rev. D* 92.9 (Nov. 5, 2015), p. 094010. DOI: 10.1103/PhysRevD.92.094010.

[5] Christian Bierlich, Gösta Gustafson, and Leif Lönnblad. "Collectivity without Plasma in Hadronic Collisions". In: (Oct. 26, 2017). DOI: 10.1016/j.physletb.2018.01.069.

[6] Christian Bierlich et al. "The Angantyr Model for Heavy-Ion Collisions in PYTHIA8". In: (June 28, 2018). DOI: 10.1007/JHEP10(2018)134.

[7] Andy Buckley et al. "Systematic Event Generator Tuning for the LHC". In: *Eur. Phys. J. C* 65.1-2 (Jan. 2010), pp. 331–357. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-009-1196-7. arXiv: 0907.2973.

[8] Ohio Supercomputer Center. *Ohio Supercomputer Center*. 1987. URL: http://osc.edu/ark:/19495/f5s1ph73.

[9] A. K. Chaudhuri. "A Short Course on Relativistic Heavy Ion Collisions". In: (July 23, 2012). arXiv: 1207.7028 [hep-ph, physics:nucl-th].

[10] ALICE Collaboration. "Higher Harmonic Anisotropic Flow Measurements of Charged Particles in Pb-Pb Collisions at 2.76 TeV". In: *Phys. Rev. Lett.* 107.3 (July 11, 2011), p. 032301. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.107.032301. arXiv: 1105.3865.

[11] ALICE Collaboration. "Long-Range Angular Correlations on the near and Away Side in p-Pb Collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV". In: *Physics Letters B* 719.1-3 (Feb. 2013), pp. 29–41. ISSN: 03702693. DOI: 10.1016/j.physletb.2013.01.012. arXiv: 1212.2001.

[12] Alice Collaboration. "Harmonic Decomposition of Two-Particle Angular Correlations in Pb-Pb Collisions at $\sqrt{s_{\rm NN}}$ = 2.76 TeV". In: (Sept. 12, 2011). DOI: 10.1016/j.physletb.2012.01.060.

[13] C. M. S. Collaboration. "Principal-Component Analysis of Two-Particle Azimuthal Correlations in PbPb and pPb Collisions at CMS". In: (Aug. 23, 2017). DOI: 10.1103/PhysRevC.96.064902.

[14] CMS Collaboration. "Observation of Long-Range near-Side Angular Correlations in Proton-Lead Collisions at the LHC". In: *Physics Letters B* 718.3 (Jan. 2013), pp. 795–814. ISSN: 03702693. DOI: 10.1016/j.physletb.2012.11.025. arXiv: 1210.5482.

[15] CMS Collaboration. "Suppression of Non-Prompt J/Psi, Prompt J/Psi, and Y(1S) in PbPb Collisions at Sqrt(sNN) = 2.76 TeV". In: *J. High Energ. Phys.* 2012.5 (May 2012), p. 63. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2012)063. arXiv: 1201.5069.

[16] Na49 Collaboration and H. Appelshaeuser et Al. "Directed and Elliptic Flow in 158 GeV/Nucleon Pb + Pb Collisions". In: (Nov. 7, 1997). DOI: 10.1103/PhysRevLett.80.4136.

[17] The ATLAS collaboration. "A Study of Different Colour Reconnection Settings for Pythia8 Generator Using Underlying Event Observables". In: *CERN Document Server* ATL-PHYS-PUB-2017-008 (May 4, 2017). URL: https://cds.cern.ch/record/2262253.

[18] A. Deur, S. J. Brodsky, and G. F. de Teramond. "The QCD Running Coupling". In: *Progress in Particle and Nuclear Physics* 90 (Sept. 2016), pp. 1–74. ISSN: 01466410. DOI: 10.1016/j.ppnp.2016.04.003. arXiv: 1604.08082.

[19] K. Dusling et al. "Long Range Two-Particle Rapidity Correlations in A+A Collisions from High Energy QCD Evolution". In: *Nuclear Physics A* 836.1-2 (May 2010), pp. 159–182. ISSN: 03759474. DOI: 10.1016/j.nuclphysa.2009.12.044. arXiv: 0911.2720.

[20] *Gnu.Org*. URL: https://www.gnu.org/software/gsl/.

[21] Ulrich W. Heinz. "Concepts of Heavy-Ion Physics". Version 1. In: (July 30, 2004). arXiv: hep-ph/0407360.

[22] Ulrich W. Heinz, Zhi Qiu, and Chun Shen. "Fluctuating Flow Angles and Anisotropic Flow Measurements". In: (Feb. 14, 2013). DOI: 10.1103/PhysRevC.87.034913.

[23] Ulrich W. Heinz and Raimond Snellings. "Collective Flow and Viscosity in Relativistic Heavy-Ion Collisions". In: (Jan. 13, 2013). DOI: 10.1146/annurev-nucl-102212-170540.

[24] M. Hippert et al. "Measuring Momentum-Dependent Flow Fluctuations in Heavy-Ion Collisions". In: *Phys. Rev. C* 101.3 (Mar. 4, 2020), p. 034903. DOI: 10.1103/PhysRevC.101.034903.

[25] Daniel Iagolnitzer, Vincent Rivasseau, and Jean Zinn-Justin. *International Conference on Theoretical Physics: TH-2002, Paris, July 22–27, 2002*. Birkhäuser, Dec. 6, 2012. 981 pp. ISBN: 978-3-0348-7907-1. Google Books: dqj5BwAAQBAJ.

[26] Avery McIntosh. "The Jackknife Estimation Method". In: (June 1, 2016). arXiv: 1606.00497.

[27] Jean-Yves Ollitrault. "Determination of the Reaction Plane in Ultrarelativistic Nuclear Collisions". In: (Mar. 11, 1993). DOI: 10.1103/PhysRevD.48.1132.

[28] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 Physics and Manual". In: *J. High Energy Phys.* 2006.05 (May 2006), pp. 026–026. ISSN: 1126-6708. DOI: 10.1088/1126-6708/2006/05/026.

[29] Carla M. Vale. "Elliptic Flow in Au+Au Collisions at RHIC". In: (Oct. 6, 2004). DOI: 10.1088/0954-3899/31/4/006.