

RICE UNIVERSITY
SENIOR THESIS

Event-by-Event analysis of Elliptical Flow

Author:

Joe Anderson

Advisor:

Dr. Wei Li

*A thesis submitted in fulfillment of the requirements
for the degree of Bachelor of Science*

in the

CMS Heavy Ion Group
Physics Department

April 22, 2013

Abstract

Lead-lead collisions are one of the current focuses of the Large Hadron Collider at CERN. It is possible that quark deconfinement occurs in such collisions, and so it is important to study the properties that manifest from the highly thermalized quarks and gluons. Flow properties play an important role in such research, and I focus on elliptic flow (v_2) in this paper. I will discuss background information as well as my results in studying the flow coefficients of simulated heavy-ion collision data. Finally, I will describe my progress in utilizing Bayesian unfolding to remove gaussian smearing from the v_2 distribution for collisions with impact parameter $8 < b < 10$.

Acknowledgements

I would like to thank Dr. Jabus Roberts and Dr. Wei Li for their time and help this year. Additionally, this work was supported in part by the Data Analysis and Visualization Cyberinfrastructure funded by NSF under grant OCI-0959097, for the procurement of DAVinCI.

Contents

Abstract	i
Acknowledgements	i
1 Introduction	1
1.1 Quark-Gluon Plasma	1
1.2 Elliptic Flow	3
Formulas	4
1.3 Bayesian Unfolding	6
2 Software	9
2.1 Simulation	9
2.2 Processing	10
Histograms	10
Trees	11
3 Results	12
3.1 Eccentricity	12
3.2 Flow Coefficients	15
3.3 Impact Parameter	15
3.4 Pseudorapidity	17
3.5 Gaussian Smearing	18
Bibliography	22

Chapter 1

Introduction

1.1 Quark-Gluon Plasma

Quark-Gluon Plasma (QGP) is a theorized state of matter in which the quarks and gluons are so energetic that they are *deconfined*.¹ This concept comes from confinement, the property that the color field between quarks increases in strength so readily with distance that pulling two bound quarks apart results in quark-antiquark pair production from the field energy. Thus, trying to break apart bound quarks results in new quarks to be bound to, and so they remain *confined*.

The theory is that high-temperature hadrons might undergo a phase transition where the quarks become deconfined. This idea follows from the lagrangian formulated in Quantum Chromodynamics (QCD), which accurately describes strong interactions and explains that gluons are self-coupling. This

¹Some papers implicitly define QGP to be the existing state of matter being studied in heavy-ion research. For that definition, QGP would be described as *possibly* consisting of deconfined quarks. I will use the definition that all of my references use, that QGP is what we hope to find, not what we have already discovered.

is what causes the aforementioned confinement, but only for small momentum transfers [1]. For large momentum-transfers, QCD predicts a state of asymptotic freedom. If such a state exists, it would be immediately imperative to investigate the properties of such a material.

Currently, it is suspected that QGP has been discovered in heavy ion research, and as such I will now discuss some historical events to date. Basic arguments understood in 1980 predicted the strong phase transition to occur for an energy density of $\epsilon \sim 1 \text{ GeV}/\text{fm}^3$ and a temperature of $T \sim 200 \text{ MeV}$, but this was well outside the range of study at the time. The 1990s mark the start of heavy ion experiments at the Brookhaven Alternating Gradient Synchrotron (AGS) and the CERN Super Proton Synchrotron (SPS). These experiments continued with the AGS serving as an injector for the Relativistic Heavy-Ion Collider (RHIC), where maximum center of mass energies of $\sqrt{s_{NN}} = 200 \text{ GeV}$ were attained, an order of magnitude higher than its predecessors [1]. Most recently, the Compact Muon Solenoid (CMS) as part of the LHC at CERN has been investigating PbPb collisions at $\sqrt{s_{NN}} = 2.76 \text{ TeV}$ [2]. These PbPb collisions are of the most interest for this paper.

Heavy-ion collisions are relevant to QGP research since they produce a relatively large amount of matter at the highest temperatures we can currently create. As of this writing, there exists significant evidence from heavy ion research that a phase transition occurs in such collisions. However, instead of the expected transition into a hot gas of non-interacting quarks and gluons, it appears that the hot quark-gluon mass behaves very much like a liquid with a viscosity to entropy ratio η/s very close to a theoretical minimum obtained using gauge gravity duality calculation [1]. This unexpected result is very interesting, but experimentally determining η/s requires an understanding of the mechanism of flow and fluidity in the hot quark-gluon matter.

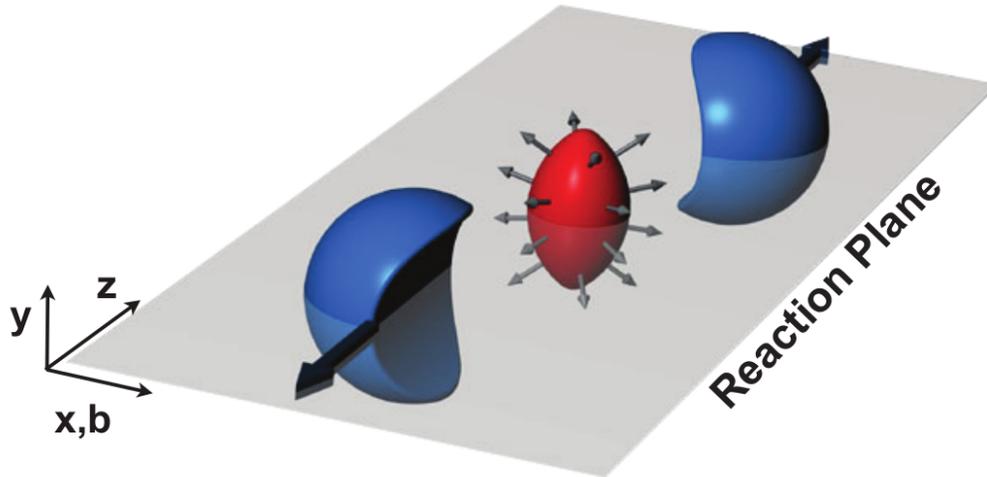


FIGURE 1.1: A 3D model of two lead ions colliding off-center, where the distance between them at the time of the collision, b , is in the x -direction [1]. Note the somewhat ellipsoid collision region.

1.2 Elliptic Flow

First, one must have an accurate picture of the sort of collision being discussed. Figure 1.1 shows a 3D model of a PbPb collision with labeled axes and reaction plane. If the hot quark-gluon ellipsoid in the middle of the collision behaves like a non-interacting gas, then it will expand uniformly in all directions. Then, as it expands and cools, the particles detected will not have any azimuthal dependence.² The actual results show that generally there is a strong increase in particle production along the short axis of the ellipsoid, which is close to the x -axis.³ A model of how this flow progresses is illustrated in Figure 1.2.

The simplest description for the flow is that initially, the fluid is compressed

²The azimuth ϕ is the angle in the xy -plane relative to the reaction plane, or counter-clockwise from the $+x$ -axis.

³See Chapter 3 for discussion of the variation in ellipsoid orientation from the reaction plane.

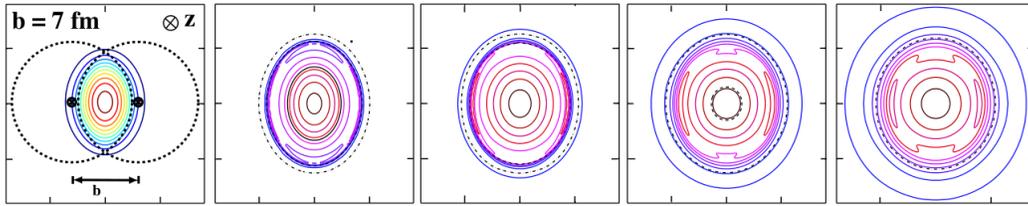


FIGURE 1.2: A profile of the energy density in the xy -plane over time for a non-central collision. The z -axis is defined to be along the incoming beam direction, and the x -axis is defined by the impact parameter b [1].

along the short axis, so like the surface tension of a water droplet, the ellipsoid “wobbles” into a spheroid as it cools. However, this is caused by a momentum transfer from the long axis to the short axis, so now what was the short axis of the ellipsoid now has a far greater momentum than the long axis. More momentum means more kinetic energy, so more particle production along the short axis is observed.

Formulas There are many ways to characterize this flow from data. To understand it from simulated data (see Section 2.1), one can look at the initial state of the participating particles and compare it to the final state of the momenta of the produced particles.

The initial state is generally characterized by using a Fourier decomposition:

$$\frac{dN}{d\phi_i} \propto 1 + 2 \sum_{n=1}^{\infty} v_n \cos n(\phi - \Phi_n^*) \quad (1.1)$$

and then finding the reference angle Φ_2 and eccentricity ϵ_2 using:

$$\Phi_n = \Phi_n^* + \frac{\pi}{n} = \frac{1}{n} [\text{atan2}(\langle r^n \sin(n\phi_i) \rangle, \langle r^n \cos(n\phi_i) \rangle) + \pi] \quad (1.2)$$

$$\epsilon_n = \frac{\sqrt{\langle r^n \cos(n\phi_i) \rangle^2 + \langle r^n \sin(n\phi_i) \rangle^2}}{\langle r^n \rangle} \quad (1.3)$$

where ϕ_i is the azimuthal angle of an initial participating particle, n is the order of the eccentricity ($n = 2$ for elliptic behavior), $\langle \alpha_i \rangle$ is the average α for the initial participating particles, and $\text{atan2}(a, b)$ is the same as $\text{atan}(a/b)$, but with the range $(-\pi, \pi]$ [2].

The final state data is similar to compute, but has more subtleties. There are now interesting issues, like how the v_n coefficient given by a Fourier decomposition of the final momenta is going to have a reference angle Ψ_n distinct from the Φ_n in the initial participant decomposition. In practice with experimental data, this is resolved by using two-particle correlations ($v_n \{2\}$) and four-particle correlations ($v_n \{4\}$) which do not depend on the reference angle, and then inferring what the mean and standard deviation of v_n is from

$$v_n \{2\}^2 = \langle v_n \rangle^2 + \sigma_{v_n}^2 \quad (1.4)$$

$$v_n \{4\}^2 = \langle v_n \rangle^2 - \sigma_{v_n}^2 \quad (1.5)$$

where the cumulants $v_n \{2\}$ and $v_n \{4\}$ involve iterating over every pair and quadruplet of particles, respectively [2].

This, however, is technically difficult, so I started with using the much simpler

$$v_n = \langle \cos(n\phi_f - n\Psi_n) \rangle \quad (1.6)$$

where ϕ_f is the final momentum azimuth and Ψ_n can either be taken to be the same as Φ_n or calculated based off the final momenta by taking Equation 1.2 but replacing $\phi_i \rightarrow \phi_f$.

1.3 Bayesian Unfolding

I warn the reader that this section is out of context no matter where it lies, so some of the bits will not make sense until that context is filled.

First, *unfolding* is the process by which one tries to solve the following problem: Given a distribution/vector of effects $v_{2\text{obs}} = n(E)$ and a smearing matrix $\Lambda = (\lambda_{ji})$ with elements $\lambda_{ji} = P(E_j|C_i)$, what is a reasonable guess for the distribution/vector of causes $v_{2\text{true}} = n(C)$?

There are many ways people have tried to solve this problem in the past. I will mention them and their advantages briefly; my references go into greater detail, so feel free to look in those documents for more information. So, the methods include [3]:

bin-by-bin Models the action of Λ as the multiplication of an efficiency factor on each bin of $v_{2\text{true}}$ to obtain directly a corresponding $v_{2\text{obs}}$.

Valid when there is negligible migration and the standard deviation is less than the bin size.

Other than being fundamentally wrong, you also have the threat that choosing smaller bins can have worse results.

matrix inversion Models the action of Λ as a deterministic matrix multiplication $v_{2\text{obs}} = \Lambda v_{2\text{true}}$. Solves for $v_{2\text{true}}$ by finding $\Lambda^{-1}v_{2\text{obs}}$.

Λ may be singular, in which case rebinning can sometimes resolve the issue.

Results are still generally unstable, in that small changes in inputs create large changes in outputs, and in that negative values can be obtained for cause distributions which should never be negative.

regularized unfolding Models the action of Λ as a composition of orthogonal polynomials.

Produces satisfactory results.

Technically difficult: only one-dimensional problems even seem solvable.

Bayesian unfolding Models the action of Λ as a set of probabilistic statements. Solves for $v_{2\text{true}}$ by iterative applications of Bayes' theorem.

Theoretically well-grounded, has no restrictions on binning.

Of the possible methods, Bayesian unfolding is a definite win. So, let's discuss the method itself. Again, I will not go into full detail for the sake of space.

The full method as described by D'Agostini[3] involves some extra steps, wherein he calculates event distributions unnecessarily, when probability distributions are perfectly effective and may be converted to event distributions at the end. A concise version says:

1. Choose an arbitrary $P_0(C)$ based on your best guess of the distribution.
2. Calculate the efficiency for each cause, given by

$$\epsilon_i = \sum_j \lambda_{ji} = \sum_j P(E_j|C_i) \quad (1.7)$$

3. Calculate the next iteration's probability distribution, given by

$$\tilde{P}_1(C_i) = \frac{1}{\epsilon_i} \sum_j P(C_i|E_j) \quad (1.8)$$

$$= \frac{1}{\epsilon_i} \sum_j \frac{P(E_j|C_i) \cdot P_0(C_i) \cdot P(E_j)}{\sum_\alpha P(E_j|C_\alpha) \cdot P_0(C_\alpha)} \quad (1.9)$$

where Equation 1.9 is given by Bayes' Theorem.

4. Normalize your distribution:

$$P_1(C_i) = \frac{\tilde{P}_1(C_i)}{\sum_{\alpha} \tilde{P}_1(C_{\alpha})} \quad (1.10)$$

5. Return to Step 3 and repeat.

Chapter 2

Software

2.1 Simulation

Since actual collision data is vast and time-consuming to sort through, before running processing code over it, I worked on code to analyse the data simulated by a program developed by RHIC researchers called “A Multi-Phase Transport model” (AMPT). There are many types of models for heavy-ion collisions; three are: thermal, hydrodynamic, and transport. The assumptions for these, respectively, are global thermal and chemical equilibrium, local thermal equilibrium, and non-equilibrium [4].

Thermal models have been good for explaining the particle yields; hydrodynamic for collective behavior of particles with low transverse (z) momentum; transport for more complicated effects like Hanbury-Brown-Twiss interferometry of hadrons. Where these fall short, perturbative QCD has produced a few models for particles with large momentum transfer, and a few special effects have developed their own models and theories, such as Yang-Mills theory [4].

All in all, for our purpose of studying the elliptic flow, hydrodynamic or transport models are suitable. However, it is far more interesting to use a more thorough transport model, so we use AMPT.

AMPT accepts a configuration file with every necessary variable about the collision, with the option to specify ranges of values. Thus, we generate a lot of data with random impact parameters b , and then sort through the data after the fact using ROOT, discussed further in Section 2.2.

One last necessary fact about our setup is that the transport model is very computationally intense, since the interactions of so many particles must be judged. As such, for data generation, we run AMPT on DAVinCI, one of the shared computing resources of the Rice Research Computing Support Group. DAVinCI was procured by the Data Analysis and Visualization Cyberinfrastructure funded by NSF under grant OCI-0959097.

2.2 Processing

I processed the data output by AMPT using ROOT. ROOT is a set of object-oriented frameworks developed by CERN and build off of C++. In addition to having fully-featured histogram, data storage, and vector methods, it features a C++ interpreter called CINT. ROOT is designed from the ground up to be a more than sufficient replacement for the FORTRAN programs that have been standard for many computationally difficult physical problems.

Histograms ROOT's histograms are versatile and streamlined for easy, rapid binning at any given moment. Additionally, plotting with any number of extra options is possible, including taking a multidimensional histogram and using one of the axes to choose colors from a colormap. Fitting is very easy as well, as illustrated in Figure 3.8.

Trees ROOT has one main data structure for compressed storage, which is the tree file. Trees are made to be very compact, but as a consequence are very nearly read-only. They are perfect for storing sets of data that all have the same overall form, such as when each collision event needs to have stored the initial positions of all participating particles as well as the final positions and momenta of all produced particles.

My advisor Dr. Wei Li and Zhenyu Chen were in charge of producing AMPT output and converting the data into ROOT tree files. Then I ran code to pass over the tree files and calculate reference angles, eccentricities, and fourier coefficients. Later, I put these values into new trees so that I could easily analyze that data for the unfolding procedure.

A large part of my research experience was learning to use C++ and ROOT for the first time. I made my own code from scratch to loop over trees and construct new ones with the analysis data. (v_n coefficients, Ψ_n measurements, etc.)

Chapter 3

Results

3.1 Eccentricity

My first task was to calculate the eccentricity ϵ_2 of the initial particles in each event using Equation 1.3.

At this point, C++ was still new to me, so I had to wrestle with many compile errors, and even a very unfortunate ROOT bug that alters warning line numbers based on the number of completed for loops in the script. However, after much effort, I finally had not only code that properly calculated ϵ_2 and Φ_2 for each event, but also ϵ_n and Φ_n for any arbitrary n . The histogram results for $n = 2$ and $n = 3$ are shown in Figure 3.1 and Figure 3.2.

It's interesting to note that the initial situation does favor noticeably larger ϵ_2 than ϵ_3 , as shown in Figure 3.1 with mean values of $\langle\epsilon_2\rangle = 0.2561$ and $\langle\epsilon_3\rangle = 0.1643$, and a wider standard deviation as well. This follows the description given in Section 1.2 that the participants should be ellipsoid initially.

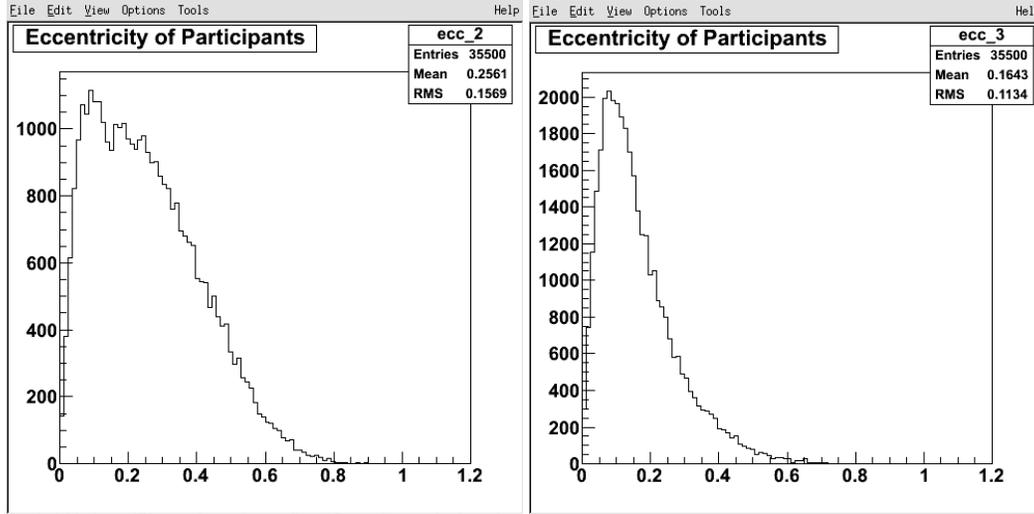


FIGURE 3.1: Eccentricity histograms for ϵ_2 and ϵ_3 for 35500 events with median bias (b randomly chosen with preference to the median value).

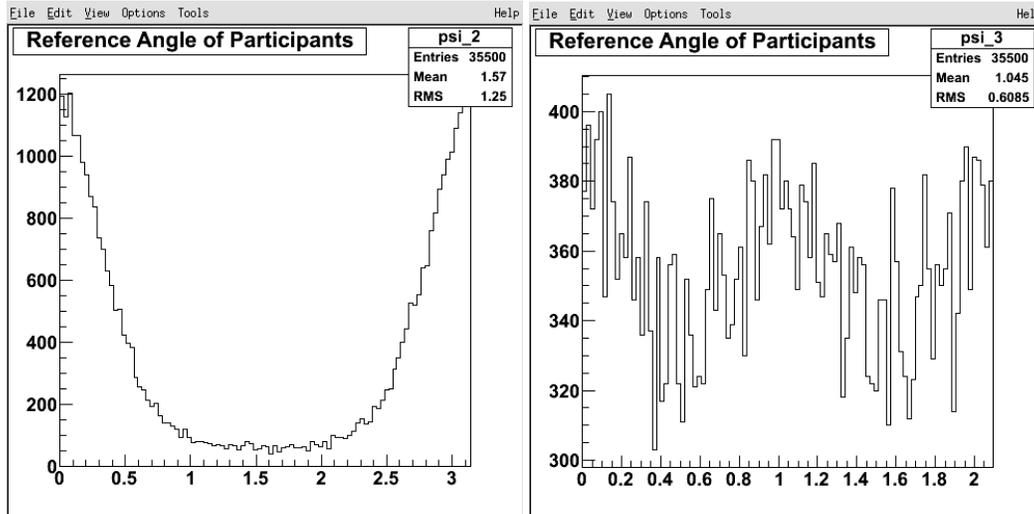


FIGURE 3.2: Reference angle histograms for Φ_2 and Φ_3 for 35500 events with median bias – at the time, I used Ψ for reference angles in general.

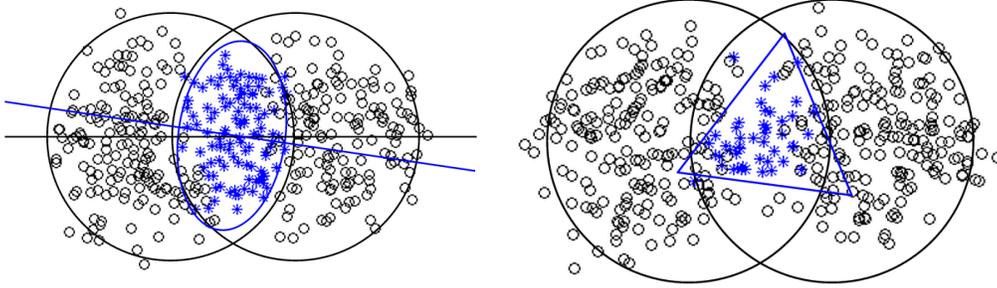


FIGURE 3.3: Initial particle position for both ions, with participating nucleons in blue. The first case shows how Φ_2 can vary from 0, and the second case shows how ϵ_3 can actually be quite large sometimes.

The other interesting note is that Φ_2 does favor $\Phi_2 = 0$ as expected, since that corresponds to the short axis of the ellipsoid being aligned with the reaction plane. The haphazard reference angles for Φ_3 indicate that there isn't a consistent phenomenon occurring there, which is also expected.

However, this begs the question: What's going on with $\Phi_2 \neq 0$ and $\epsilon_3 \gg 0$? To answer this question, I searched the median bias event simulations for cases that matched the unusual situations I was looking for, then wrote a ROOT script to plot the positions of the initial particles in those cases, shown in Figure 3.3.

These results suggest what's really happening is just due to the fact that even though lead nuclei are rather large, they still have a finite number of particles, and they aren't distributed evenly, so there can be unusual gaps in the region of intersection for the ions where no nucleons are there to represent it. Additionally, some nucleons can be in the region of intersection where there are plenty of nucleons, but still happen to pass through without interaction. So, the ellipsoid-like shape in Figure 1.1 is very much an idealistic picture, the product of perfectly uniform spheres of nucleons colliding with each other.

3.2 Flow Coefficients

With a firm understanding of the initial conditions, I could then start investigating the flow coefficients v_n , given by Equation 1.6.

Calculating v_n for arbitrary n was easy, now that ROOT and C++ were starting to become familiar. However, it was clear early on that some focus was necessary, and under Wei Li's advice, I started coding specifically for second-order flow.

Thus, I began searching for an expected correlation between v_2 and ϵ_2 . This proved fruitful, but not completely clear. Wei further advised me to start selecting which events I was analyzing as oppose to taking in all median bias events. Additionally, once it became clear that v_2 was behaving as expected with respect to ϵ_2 , we moved past studying the initial particle state entirely, since it's not observable for actual data, anyway.

To segue into the next section, a v_2 distribution for median bias, $8 \leq b \leq 10$, is shown in Figure 3.4

3.3 Impact Parameter

At this point, choices of impact parameter had to be made. I would like to take a moment to discuss the implications of this. I'm starting to focus on data that would show up in actual experimental data, not just simulated data, and yet I'm going to be specially selecting impact parameter? Surely, we cannot measure the separation distance between two relativistic heavy ions on the order of a few fermi.

However, it is actually very reasonable to make this decision, as we have an indirect method for measuring impact parameter. The related quantity is

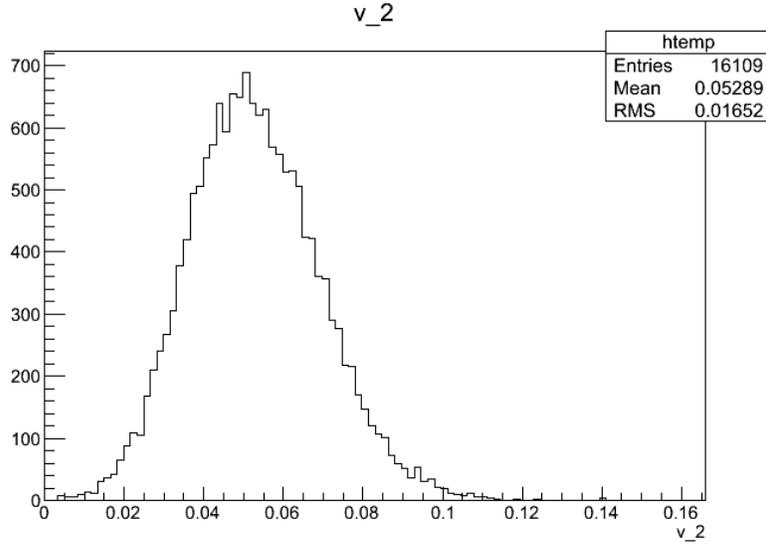


FIGURE 3.4: A histogram of v_2 values for a relatively small number of median bias events with impact parameter $8 \leq b \leq 10$.

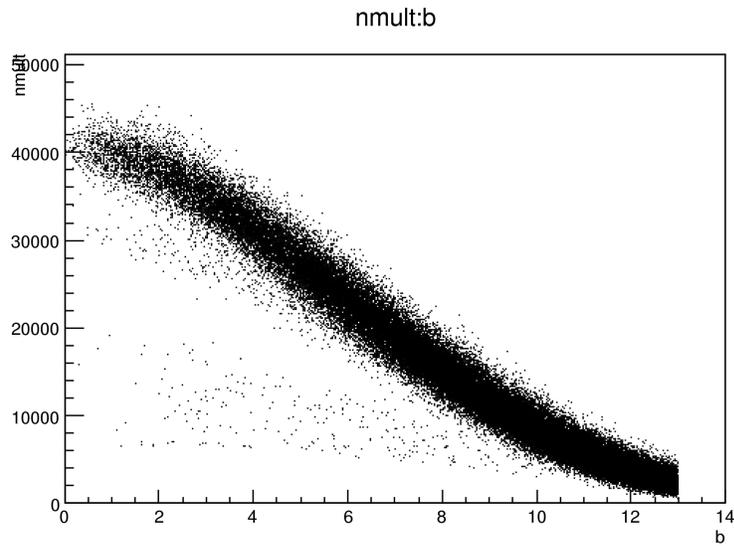


FIGURE 3.5: A scatterplot of the number of produced particles N_{mult} versus the impact parameter b (in fermi).

overall number of final particles measured, N_{mult} . It makes sense for N_{mult} to be negatively correlated with b , since $b = 0$ means the ions are colliding head-on, so this should be the case where the greatest number of particles N_{mult} are produced. Interestingly, the correlation is not only consistent, but very reliably so; this partly comes from the fact that the maximum N_{mult} is very large, on the order of tens of thousands, as shown in Figure 3.5's scatterplot of N_{mult} versus b . In fact, Figure 3.5 makes it rather clear that knowing N_{mult} means you can feel rather confident you know b to within ± 1 fermi.

So, it is very reasonable to select a region of b -values that is 2 fm wide and expect it to be experimentally determinable as well. In particular, we select $8 \leq b \leq 10$ so that the intersection has a good chance of being elliptic and $N_{mult} \sim 10000$.

3.4 Pseudorapidity

Another way one can break up the data for interesting research is by the pseudorapidity η . This is a spatial coordinate, not an event property, and it represents the angle of a particle relative to the beam axis:

$$\eta = -\ln \tan \frac{\theta}{2} = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_T}{|\mathbf{p}| - p_T} \right) \quad (3.1)$$

where θ is the angle relative to beam and p_T is the transverse momentum p_z .

How this can be used in analysis is by studying the v_2 and Ψ_2 values when you only take certain η -ranges into account. The simplest of these are the ranges $\eta > 0 \iff p_z > 0$ and $\eta < 0 \iff p_z < 0$. These each take half of the particles into account, and in general, should be symmetric. That is, the same v_2 coefficient should be obtained for both sets of particles, since v_2

is a measure of the azimuthal distribution in the xy -plane, and the division $\eta = 0$ is the xy -plane. Thus, the ranges of positive η versus negative η are symmetric about x and y , since it's merely stating whether p_z is positive or negative.

Where this becomes really powerful is that if, in one event, you measure v_2^+ as the coefficient taking only positive η into account and v_2^- as the coefficient taking only negative η into account, the absolute difference $v_{2\text{diff}} = \frac{1}{\sqrt{2}} |v_2^+ - v_2^-|$ is a direct measure of the systemic error due to N_{mult} being finite. In other words, The error due to finite particle production also causes asymmetries with respect to η , and $v_{2\text{diff}}$ is a measure of that systemic error.

3.5 Gaussian Smearing

At this point, I wrote a script to compute v_2^\pm and $v_{2\text{diff}}$. The results from a relatively small number of median bias events are shown in Figure 3.6 and Figure 3.7. It is readily apparent that some sort of gaussian-like behavior is occurring for $v_{2\text{diff}}$. To confirm this, I took in all of the median bias data we had collected thus far, still selecting $8 \leq b \leq 10$, and fit the $v_{2\text{diff}}$ distribution to a gaussian. The result of this is shown in Figure 3.8.

From here, I would like to take advantage of having found my systemic smearing distribution. This process is called *unfolding*, and in particular, the one I used is called Bayesian unfolding.

To apply Bayesian unfolding, I needed a smearing matrix of $\lambda_{ji} = P(E_j|C_i)$, where the E_j are effect bins of possible values to measure for $v_{2\text{obs}}$ and C_i are cause bins of the actual values for $v_{2\text{true}}$. To build the matrix, I took the histogram of Figure 3.8 and filled each column for a particular cause $P(E_j|C)$ with the bins from Figure 3.8, centered at the $i = j$ bin. This has a few technicalities. For small i, j , there is some confusion about running into

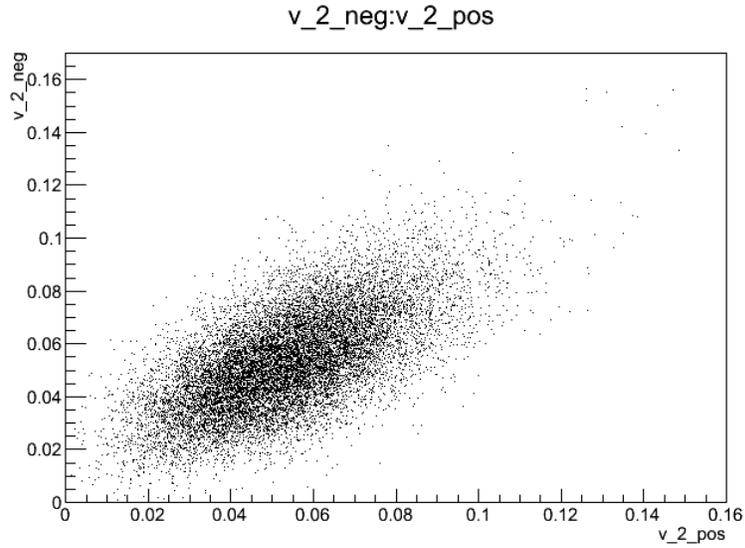


FIGURE 3.6: A scatterplot of v_2^- versus v_2^+ , illustrating the positive correlation about $v_2^+ = v_2^-$ and the deviations from this norm.

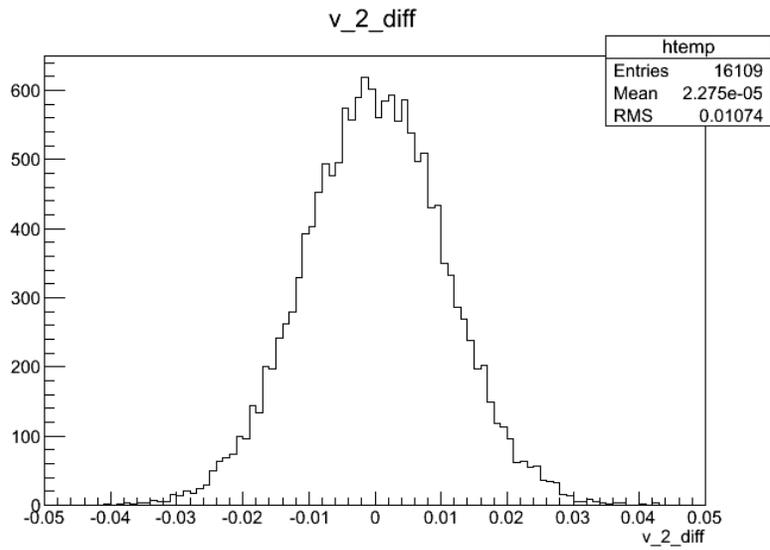
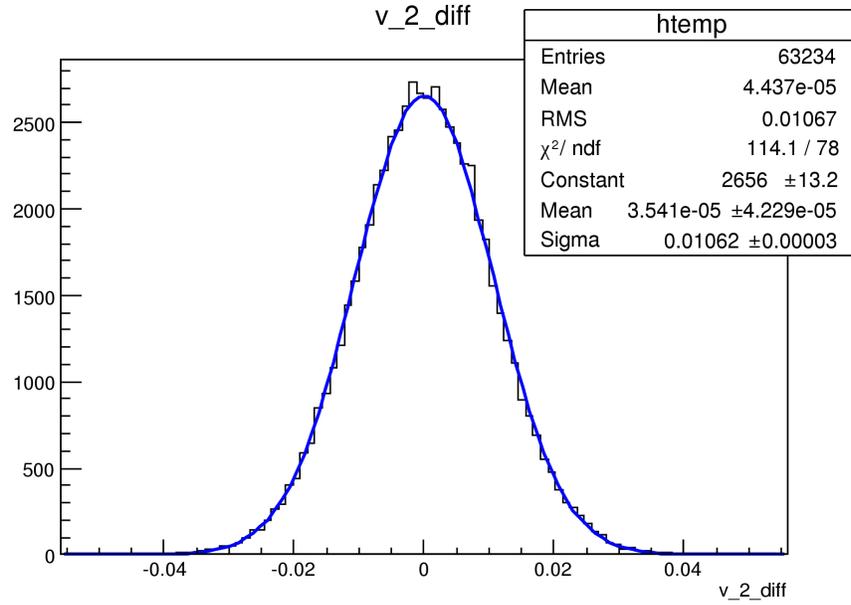
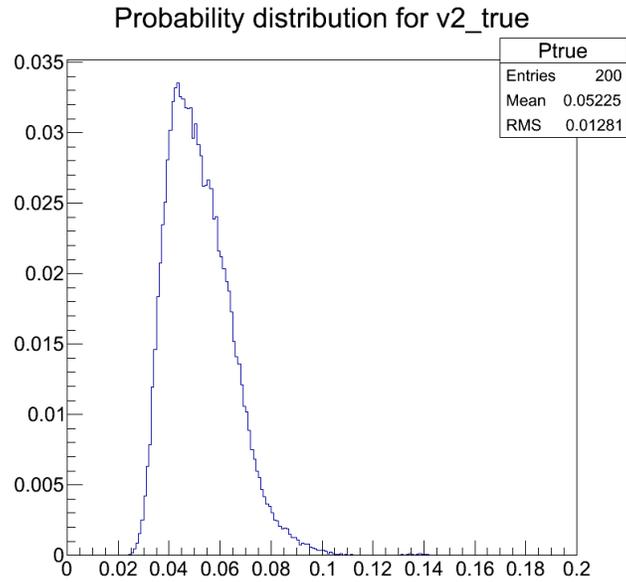


FIGURE 3.7: A histogram of $v_{2\text{diff}}$, illustrating the gaussian-like behavior and the mean very close to 0.

FIGURE 3.8: A histogram of $v_{2\text{diff}}$ with many events fit with a gaussian.FIGURE 3.9: A histogram of $v_{2\text{true}}$ with many events unfolded using $v_{2\text{diff}}$.

the edge with the gaussian histogram. The solution I used was to reflect the negative bins back into the matrix so that the C_i for small i still have full efficiency (see Equation 1.7). For the other corner issue, when i, j are large, I simply truncated the gaussian. The efficiency on that corner is irrelevant, since my choice of binning for $v_{2\text{true}}$ and $v_{2\text{obs}}$ makes those C_i have a negligible number of events.

With smearing matrix in hand, I proceeded to write a script to unfold the $v_{2\text{obs}}$ distribution I obtained from the same 63234 events I used to make Figure 3.8. The resulting $v_{2\text{true}}$ distribution is shown in Figure 3.9.

The results were only mildly satisfying. The algorithm did not appear to want to converge, possibly because of a messy Λ matrix. Further investigation into how things change if I use the gaussian I fit to Figure 3.8 to build the matrix, if I use one of the online packages for Bayesian unfolding, or if I build my smearing matrix the same way as in the ATLAS[2] paper are all good directions to continue my research.

Further, after my code runs properly on simulated data, it may be run on experimental data from the LHC.

Bibliography

- [1] Raimond Snellings. Elliptic flow: A brief review. 2011. URL <http://ezproxy.rice.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsarx&AN=1102.3010&site=eds-live&scope=site>.
- [2] Measurement of the distributions of event-by-event flow harmonics in pb-pb collisions at $\sqrt{s_{NN}}=2.76$ tev. Technical Report ATLAS-CONF-2012-114, CERN, Geneva, Aug 2012.
- [3] D'Agostini, G. A multidimensional unfolding method based on bayes' theorem. *Nuclear Inst. and Methods in Physics Research, A*, 362(2-3):487–498, 1995. ISSN 01689002. URL <http://ezproxy.rice.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-0000496635&site=eds-live&scope=site>.
- [4] 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. doi: 10.1103/PhysRevC.72.064901.