

# A Study of Charm Quark Correlations

PHY 371C: Presentation

---

Dev Desai

# Table of contents

1. Introduction
2. Theory Background
3. Pythia
4. Phase Space Cut Analysis
5. Charm Multiplicity Analysis
6. Kinematic Correlations
7. Next Steps
8. Thank you for listening! Any questions?

# Introduction

---

Why study charm hadrons?

- Charm quarks are created almost exclusively in the initial hard scattering
- Allows us to probe QCD collision dynamics.

How do the kinematic distributions of charm hadrons produced in high-energy collisions reflect the underlying charm production mechanisms?

# Goals

- Read and learn the basics of particle physics.
- Familiarize myself with Monte Carlo event generators like Pythia, and understand their limitations
- Explore the various factors affecting collisions, such as  $\sqrt{s}$ ,  $p_T$  Hard, and particle filtering, and analyze their effects on charm quark production and kinematic properties
- Develop a strong foundation and passion for particle physics data analysis and theory
- Publish a GitHub repo to accelerate future undergraduate research

# Theory Background

---

# Hard Scattering and Hadronization

## Hard Scattering:

- High-energy parton-parton collisions (e.g.,  $gg \rightarrow c\bar{c}$ )
- Sets the initial kinematic scale of the event

## Hadronization:

- Partons (quarks, gluons) cannot exist freely
- They transform into hadrons (e.g.,  $c \rightarrow D^0, \Lambda_c$ )



# Transverse Momentum ( $p_T$ )

$$p_T = \sqrt{p_x^2 + p_y^2}$$

- Momentum in the plane **perpendicular** to the beam
- Collisions at the LHC happen along the z-axis —  $p_T$  is what we actually measure
- High  $p_T$  implies a more energetic process

# Azimuthal Angle ( $\phi$ )

- Angle in the transverse (x-y) plane
- Ranges from  $-\pi$  to  $\pi$
- Used in angular correlations (e.g.,  $\Delta\phi$  between hadrons)
- Distribution is usually flat  $\rightarrow$  no preferred direction

# Rapidity ( $y$ )

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right)$$

- Measures how forward or backward a particle goes
- Useful for massive particles (like charm hadrons)
- Boost-invariant along z-axis — great for comparing frames

# Pseudorapidity ( $\eta$ )

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]$$

- Approximation of rapidity for **massless or high-energy** particles
- Defined purely by the angle  $\theta$  from the beamline
- Detectors use  $\eta$  because they measure angle, not energy directly

## Why Use $\eta$ and $y$ , Not $\theta$ ?

- Detectors are cylindrical, aligned along the beam (z-axis)
- Measuring the polar angle  $\theta$  precisely is hard
- $\eta$  and  $y$  are designed to work well with the detector geometry and boost invariance

# Pythia

---

# What is Pythia?

**Pythia** is a Monte Carlo event generator used to simulate high-energy particle collisions, especially in proton-proton environments like those at the LHC.

- Models the full evolution of an event:
  - **Hard scattering**: primary parton interaction (e.g., charm quark production).
  - **Parton showering**: emission of gluons and quarks (initial/final state radiation).
  - **Hadronization**: transition of partons into observable hadrons.
  - **Decay**: of unstable hadrons into final-state particles.
- Highly configurable: allows tuning of energy, cuts (like **pTHatMin**), and filtering for specific processes or particles.

*Used extensively in both theory and experimental workflows to simulate detector-level events and explore particle production dynamics.*

How do we specifically need to use Pythia?

- We wish to study charm quarks, which are sparsely produced.
- We then need to analyze many events to gain a useful amount of data
- Pythia is a simulator, but we don't want to run simulations each time we do an analysis
- → Need a data structure which stores all event info, including kinematics, production vertex, and mother daughter relations



## Three possibilities

1. Find a built in Pythia command, reminiscent of Ali-physics analysis
2. Use hepMC, a general high energy physics data structure, which is consistent throughout all event generators
3. Build our own data structure

I experimented with both Python and C++ scripts and documentation for each option

# What didn't work

## 1. No built in commands available through Pythia

## 2. hepMC

- Various file formats available, such as default ASCII, binary, ROOT, and ROOT trees
- hepMC also has built in parsing methods to make analysis easy
  - ASCII (Python & C++), works, but incredible large file size.
  - Binary (only C++)
  - ROOT and ROOT Trees (only C++), ran into many dependency issues
- Overall, hepMC could work, but is difficult to set up and requires C++ if we want to store a significant number of events

# What did work

Created my own data structure: created custom ROOT tree, which branched on TParticle (ROOT class which stores particle information).

## Benefits

- Entirely based on Python
- Easy setup with Conda environments
- Uses TParticle, which makes for easy querying of the data structure for any kinematics or decay chain information you may need
- Compact file size: ~20% the size of hepMC ASCII format

# Limitations of Pythia

As I was testing the various data structures, I wanted to ensure that each was storing the same information.

**Two tests:** PDG (particle identifier) counts, and pion mothers' PDG counts.

- Found unphysical artifacts (PDG of 90, which isn't a particle) which are likely intermediate internal Pythia objects in my data structure.
- Found an extremely high number of gluons (PDG of 21) in the hepMC ASCII
- Inconsistent, but similar number of each type of quarks between the two

# Limitations of Pythia

What was the reason for these discrepancies?

→ Pythia is *not* accurate to the quark/gluon level.

**Solution?**

→ Filter out all quarks and gluons during analysis

*When I ran both tests with this filtering, there weren't any more discrepancies. Yay! We can now move onto physics. Oh, and by the way, I am analyzing proton-proton collisions*

# Phase Space Cut Analysis

---

# Default Charm Production Frequency

We wish to analyze charmed hadrons to understand charm production mechanisms. Hence we need a sizeable dataset of charm quarks.

But, if we run default Pythia events, only 3.3% of events have charm quarks.

→ Most of our data set is unusable to us.

→ Inefficient simulation and analysis

# Default Charm Production Frequency

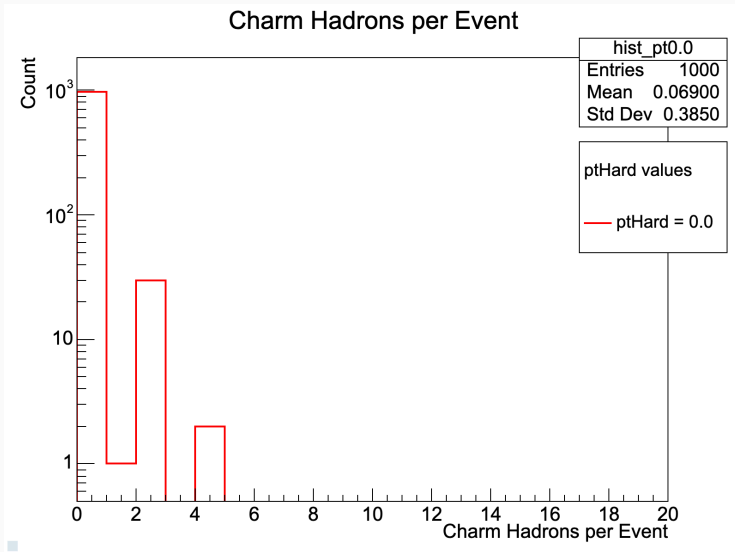


Figure 1: Histogram of charm production frequency



## How can we get a dataset which contains more charmed hadrons?

- Charmed hadrons, because of their higher mass and energy, are produced in more energetic hard scattering events.
- We can quantify with  $p_T$  Hard - how much transverse momentum is in the scattering event
- We can implement *phase space cuts* (using pTHatMin) - as Pythia is generating an event, if the  $p_T$ Hard is not  $> p_{THatMin}$ , it rejects the event and doesn't store it

# pTHatMin Analysis

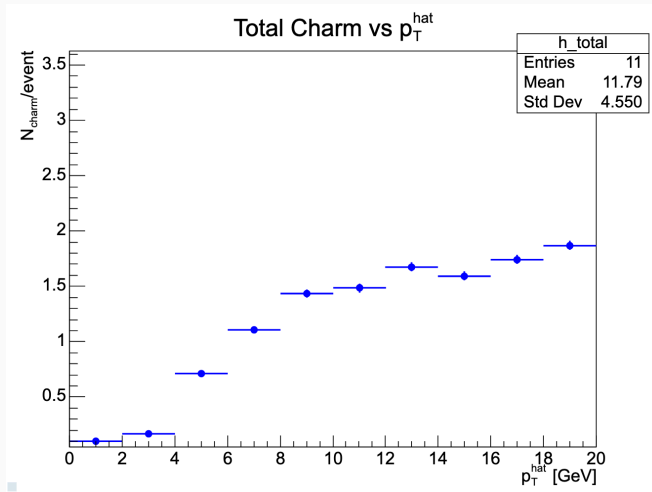


Figure 2: Phase space cuts' relation to charm production. **Note:** after  $p_T^{\text{hatMin}} = 10$ , starts to plateau

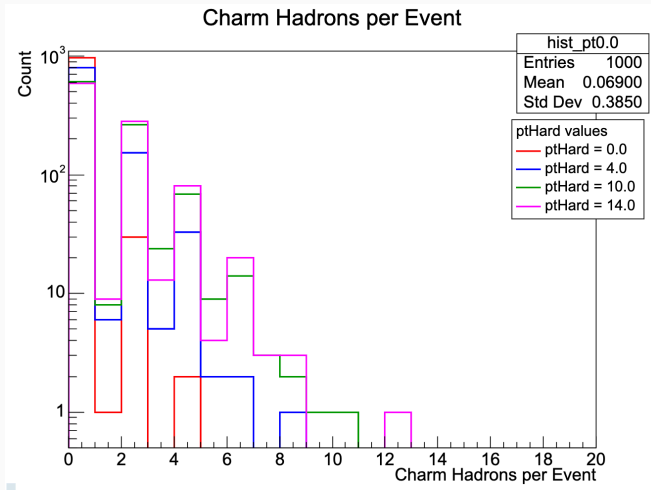
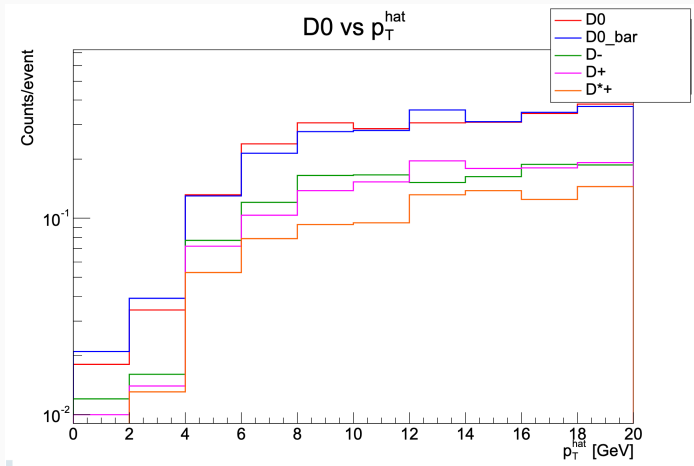


Figure 3: More context behind pTHatMin vs charm hadron production.



**Figure 4:** D Meson production with respect to pThatMin. **Note:** D0 meson's are produced at higher rates than other D mesons

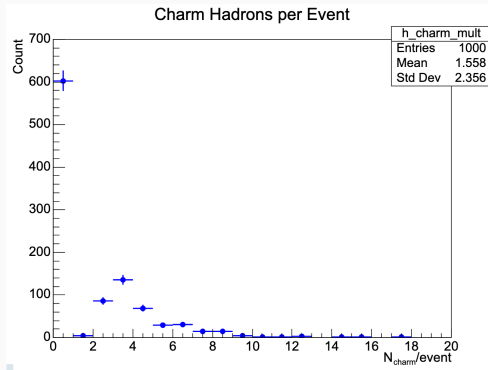
- As we increase pTHatMin, Pythia simulations take much longer, as it is rejecting more and more events because they do not meet the pT Hard cut.
- **I chose to use pTHatMin = 10 GeV for all future work**, as there was not any considerable gains after.

→ Next, I simulated 10,000 events, which took a couple of minutes, but for future work, 1 million events is preferable. This will take a significant time to simulate, and the diminishing charm production gains after pTHatMin of 10 GeV might be not worth the extra time when running the simulation.

# Charm Multiplicity Analysis

---

# Charmed Hadron Count



**Figure 5:** A histogram of charmed hadron production per event

We see something odd here: charm quarks, and thereby hadrons, are most commonly produced in pairs. But, we see that there are more events with 3 charm hadrons than 2!

# Charm Production Processes

**Pair Production:**  $gg \rightarrow cc$  or  $qq \rightarrow cc$

- Produces a pair charm quarks  $\rightarrow$  2 charm hadrons (leading to even count)
- Is the most common production mechanism

**Flavor Excitation:**  $gc \rightarrow gc$  or  $qc \rightarrow qc$

- Produces 1 charm quark  $\rightarrow$  1 charm hadron (leading to odd count)
- Rarely happens

**Gluon Splitting:**  $g \rightarrow cc$

- Produces a pair charm quarks  $\rightarrow$  2 charm hadrons (leading to even count)

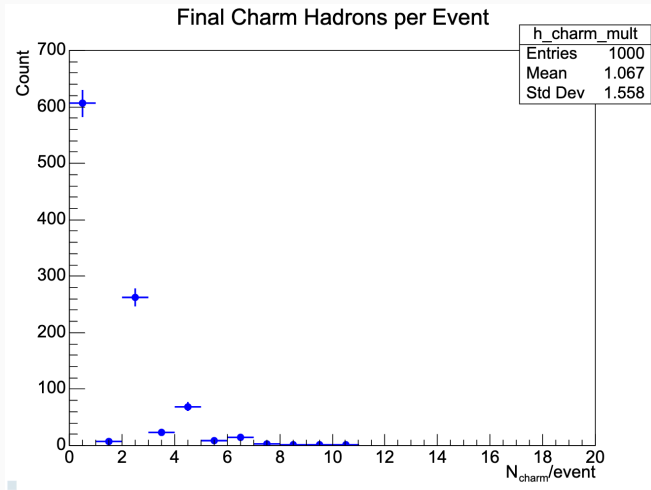


# Charm Production Processes

Charm hadrons are most commonly formed in pairs, so why the peak at 3 per event?

- When  $D^*$  mesons are produced, they then decay into a D meson. So, by counting all charmed hadrons, including higher energy states such as  $D^*$ , we are *double counting*.
- To alleviate this, instead of filtering for all charmed hadrons, we can filter for "final" charm hadrons, where they won't decay into another charm hadron.
- These include  $D^0, \overline{D}^0, D^+, D^-, D_s^+, D_s^-, \Lambda_c^+, \Lambda_c^-$

# Accounting for $D^*$ Decay



**Figure 6:** Charmed "final" hadron counts. Clear pair bias now. Still a few odd events which correlates to the rarity of flavor excitation

# Charm Counts Probability Distribution

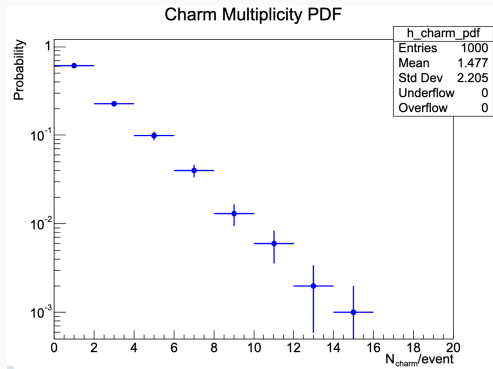
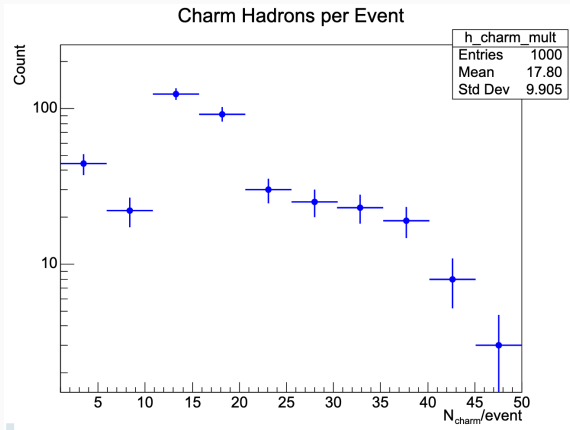


Figure 7: Charm production counts with Bin Width of 2

- Integrating from 2 to  $\infty$  yields  $P(N_c \geq 2) = 0.3880 \pm 0.0247$
- Note the exponential behavior

# Charm Quark Multiplicity



**Figure 8:** I experimented by adding in the multiplicity of charm hadrons + charm quarks to the production histogram. **Note:** there are many charm quarks counted

There are many charm quarks counted. This is misleading though, because we are parsing through all particles through the decay chain.

If we count a charmed hadron, we also counted all the vertexes previously with its charm quark. A similar pattern emerges due to jet fragmentation.

→ Leads to massive double counting.

# Kinematic Correlations

---

- When doing kinematic analysis, 1,000 events was not enough. The plots were too noisy, so I worked with 10,000 events.
- We will conduct analysis on the "final" charm hadrons.
- I also did kinematic analysis only on events which contained *only* two charm quarks, as we expect these to be correlated.

We will look at  $pT$ ,  $y$ ,  $\eta$ ,  $\phi$ , and  $\Delta\phi$  of the charmed hadrons

# Transverse momentum

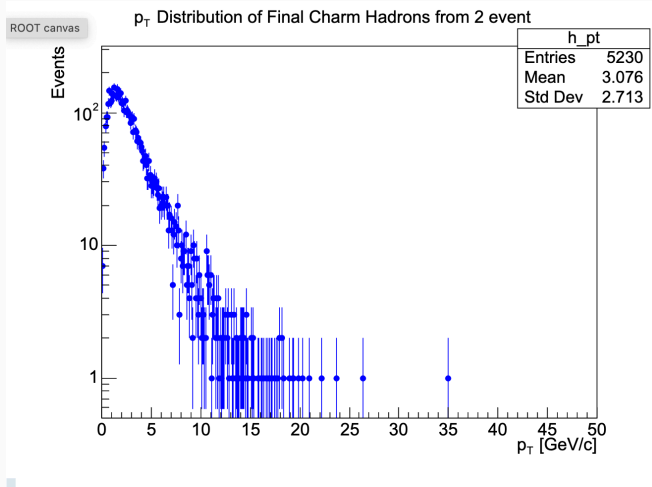


Figure 9: The  $p_T$  distribution for charmed hadrons



- It is low near  $p_T = 0$  GeV/c  $\rightarrow$  Charm quarks are produced in hard QCD processes, where some momentum transfer is required, so it is kinematically suppressed.
- There is a peak at  $\sim 2$  GeV, indicating the usual range for charmed hadrons.
- The drop off after 2 GeV, especially with the log scale, indicates the rarity of higher  $p_T$  charm hadrons.

# Rapidity and Pseudorapidity

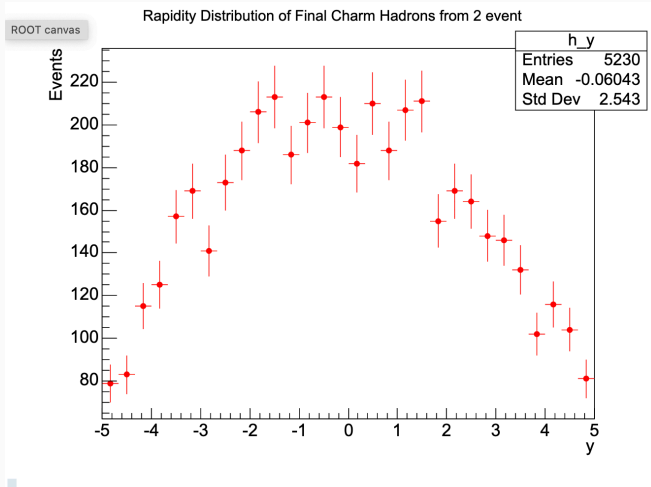


Figure 10: Rapidity distribution. **Note:** The peak at 0

# Rapidity and Pseudorapidity

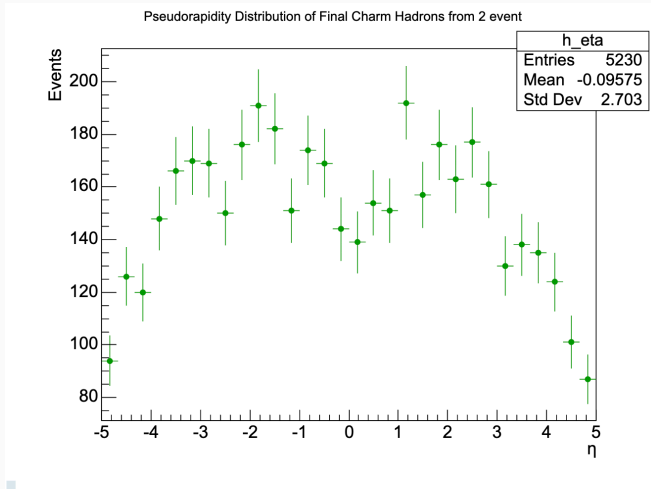
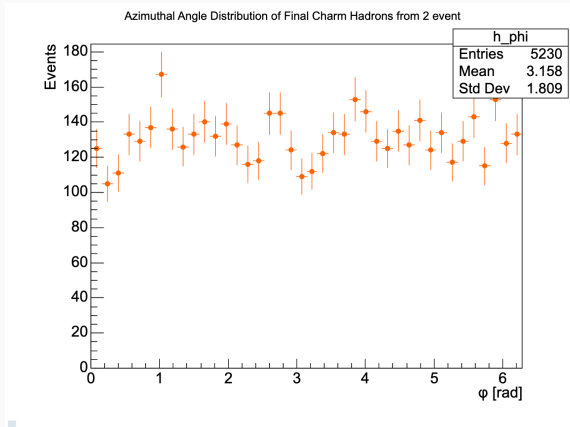


Figure 11: Pseudoapidity distribution. **Note:** The dip at 0

These distributions relate to the polar angle with respect to the beamline

- On average, the particles are centered around the pT plane with  $y = 0$ . (Polar angle of 90)
- Fewer particles are going in the direction of the beamline (very positive or negative values)
- There is a dip in  $\eta$  at 0 due to the way it is calculated.



**Figure 12:** Phi distribution. **Note:** It is randomly spread

The random spread indicates that there is no bias towards a specific direction in the pT plane.

# Delta Phi: Hadrons

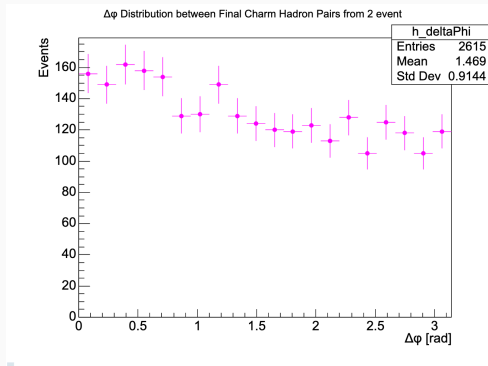


Figure 13:  $\Delta\phi$  distribution. **Note:** It is randomly spread

This is concerning - we expect there to be strong correlations, especially because we only picked events which have 2 charmed hadrons which should be correlated. We expect a peak at 0 and a broad peak on the away side.

## Including all events

I conducted the same analysis without filtering for 2 events. and got similar results for all the kinematics.

# Transverse momentum

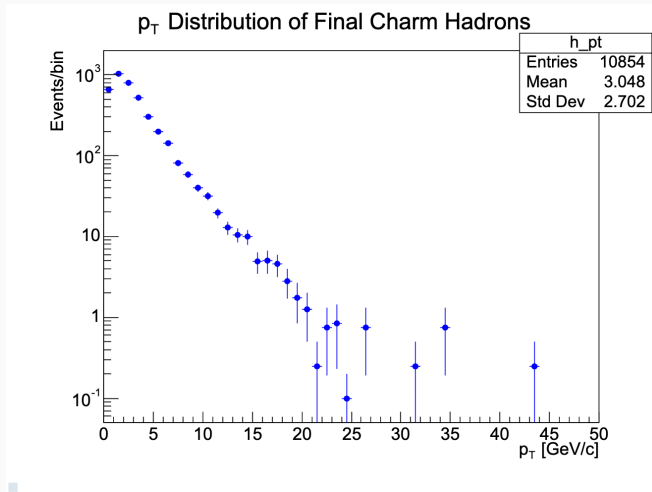


Figure 14: The  $p_T$  distribution for charmed hadrons



# Rapidity and Pseudorapidity

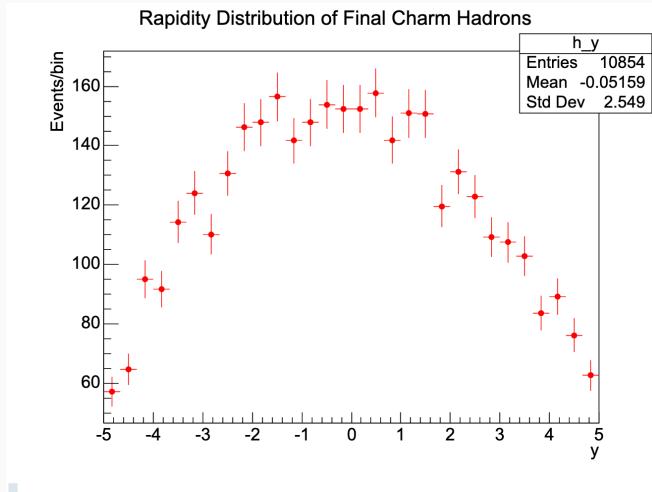


Figure 15: Rapidity distribution. **Note:** The peak at 0

# Rapidity and Pseudorapidity

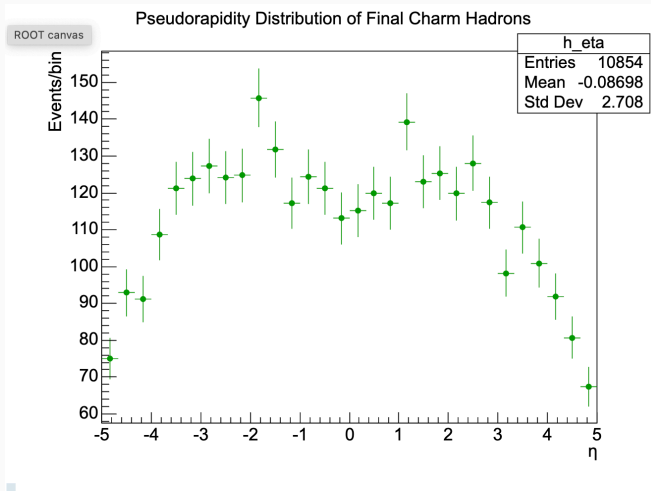
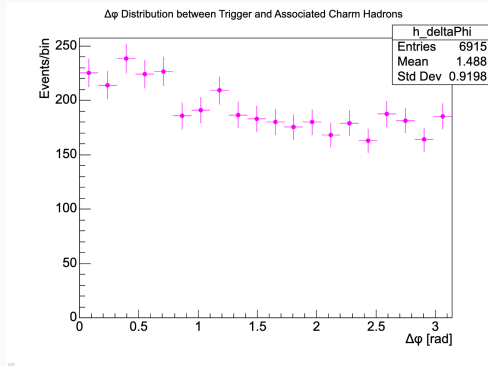


Figure 16: Pseudoapidity distribution. **Note:** The dip at 0

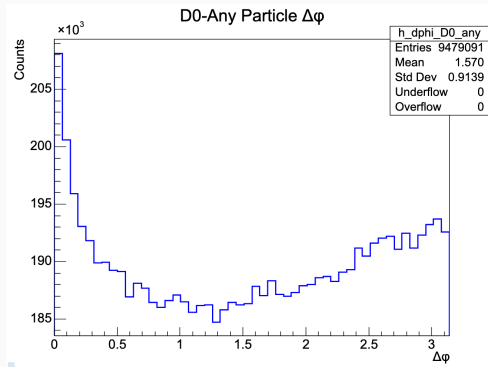
# Delta Phi: D0 Trigger



**Figure 17:**  $\Delta\Phi$  distribution. **Note:** Triggered on D0, with associated particles being all other charmed hadrons

We still see the same concerning behavior.

# Delta Phi: D0 Trigger with all other particles



**Figure 18:**  $\Delta\Phi$  distribution. **Note:** Triggered on D0, with associated particles being all other particles

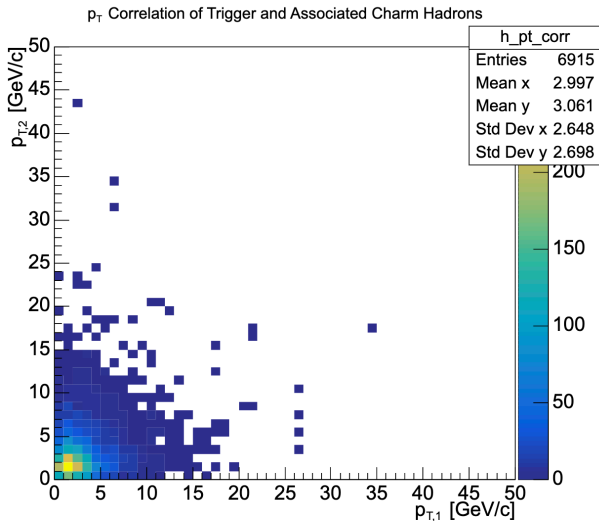
This shows the correlation which we want - a peak at 0 and a broad peak away.

## Delta Phi: D0 Trigger with all other particles

**Sharp peak at 0**  $\rightarrow$  in gluon splitting, there is 1 gluon which splits into 2 charm quarks. These quarks are close to collinear, leading to front to front peak being sharp

**Broad peak at  $\pi$**   $\rightarrow$  in pair production, there is 2 gluons which become 2 charm quarks. These quarks are opposite in phi, leading to back to back correlations. Due to fragmentation and other interactions, the peak smears.

To revisualize the data, we can use a 2D heatmap and trigger on a D0. (Trigger on D0 because it is produced the most)

Figure 19: 2D  $p_T$  Correlation.

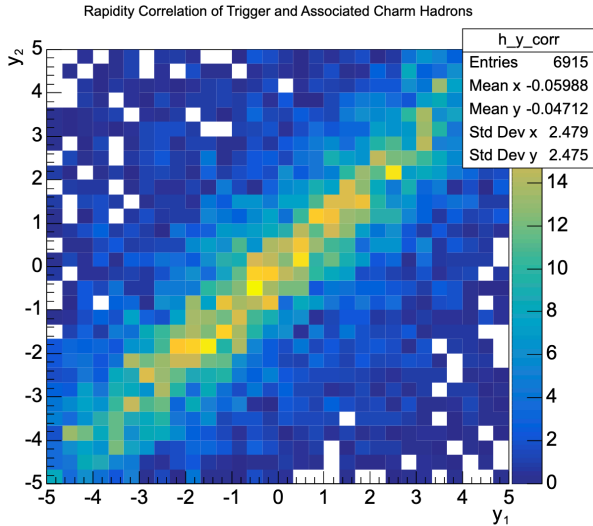


Figure 20: 2D y Correlation.



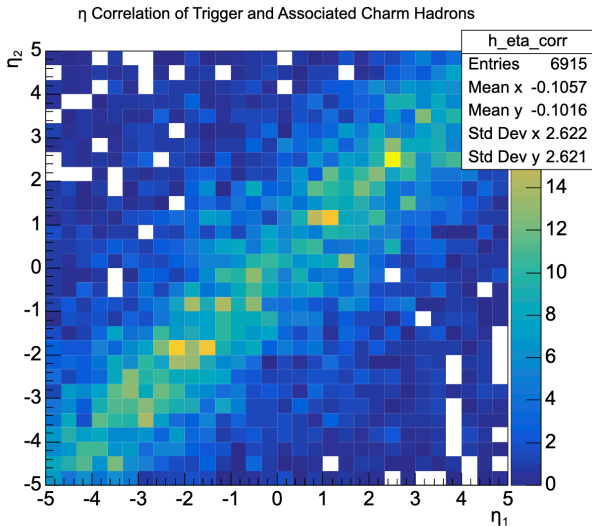


Figure 21: 2D  $\eta$  Correlation.

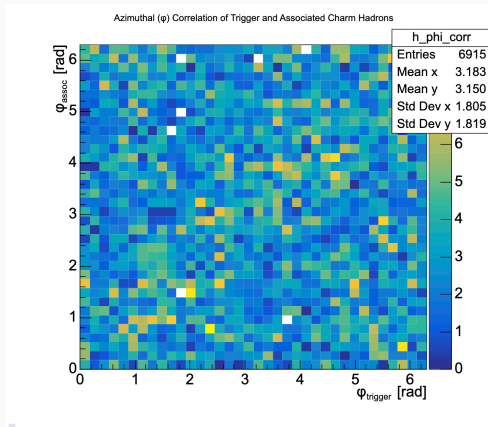


Figure 22: 2D  $\phi$  Correlation.

We can see the random behavior, similar to the  $\Delta\Phi$  plot before. Why this occurs, I don't know yet. Cliffhanger!

## Next Steps

---

# Next Steps

1. Probe the reason for the odd behavior for  $\Delta\phi$  more.
2. Publish my repository to GitHub
3. Return to write my thesis!

Thank you for listening! Any  
questions?

---