

Direct Photon Identification in Heavy Ion Collisions at sPHENIX

Chase Roumain Smith

Heavy Ions Group
Department of Physics
University of Colorado at Boulder
April 3rd, 2019

Honors Committee Members

Research Advisor: Dennis Perepelitsa - Department of Physics

Honors Council Representative: John Cumalat - Department of Physics

Committee Member: Dirk Grunwald - Department of Computer Science

Contents

1	High Energy Collisions	2
1.1	Proton Proton Collisions	2
1.2	Hard Subprocess	2
1.3	Jets	4
1.4	Photon Jet Pairs	6
2	Heavy Ion Collisions	7
2.1	Heavy Ion Collisions	7
2.2	Quark Gluon Plasma	8
2.3	Jet Quenching	9
2.4	Background	9
3	sPHENIX	11
3.1	Useful Values	11
3.2	RHIC and PHENIX	12
3.3	sPHENIX	13
3.4	Clusters	15
4	Simulation	16
4.1	Generating Events	16
4.2	Gold Gold Background	16
4.3	Detector Simulation	16
5	Cluster Isolation	18
5.1	Background Signals	18
5.2	Goals	19
5.3	Testing	20
5.4	Moving forward	22
6	Machine Learning	23
6.1	What is Machine Learning?	23
6.2	Toolkit for MultiVariate Analysis (TMVA)	23
6.3	MLP	24
6.4	BDT	24
6.5	kNN	25
6.6	Likelihood	26

7	Identifying Direct Photons	27
7.1	PHENIX Algorithm	27
7.2	Goals	28
7.3	Challenges	28
7.4	Training 1	29
7.5	Training 2	30
7.6	Moving Forward	32
8	Conclusion	33

List of Figures

1.1	Example of two quarks interacting strongly during a proton proton collision.	3
1.2	Two quarks interacting and producing two quarks (a), a quark and a gluon interacting to produce a quark and a gluon (b), two gluons interacting and producing a quark and antiquark pair (c).	3
1.3	Quark (Blue) travelling into page, quark (orange) travelling out of page.	4
1.4	A quark exiting the hard subprocess emitting a gluon, and creating a shower of quarks and gluons that will become a jet.	4
1.5	Two jets exiting a proton proton collision, known as a dijet event . .	5
1.6	Two gluons interacting to form a photon and a gluon (a), a quark and gluon interacting to form a quark and photon (b).	6
1.7	A photon jet pair from a proton proton collision.	6
2.1	The history of the universe as told from a particle point of view. . . .	7
2.2	QCD Phase diagram showing states of matter and where different experiments probe the diagram.	8
2.3	Jets being quenched differently depending on the distance they travel through quark gluon plasma.	9
3.1	Diagram showing different values of pseudorapidity.	11
3.2	Aerial Photograph of the Relativistic Heavy Ion Collider	12
3.3	Cross section view along the beam line of PHENIX and the detector that compose each of its arms [1].	12
3.4	CAD rendering of sPHENIX and the detectors that compose each of its layers [1].	13
3.5	Slice of the sPHENIX calorimeters and magnet showing opposite alignment of inner and outer HCal towers.	14
4.1	The flow of computer simulation. Pythia8 events are produced and along with already simulated gold gold event are put into a full detector simulation which then outputs truth and reconstructed information about the event.	17
5.1	A side view of the sPHENIX tracker hits (purple dots), trajectories in the tracker (black), electromagnetic calorimeter (blue), inner hadronic calorimeter (green), and outer hadronic calorimeter (red).	19
5.2	Transverse isolation energy of direct photon clusters and decay photon clusters using a cone size of $R = 0.3$	20

5.3	Cartoon showing in principle the background subtraction that is run on gold gold events.	20
5.4	Transverse isolation energy of direct photon clusters calculated using cone sizes of 0.2, 0.3, 0.4. Photon jet pairs generated in Pythia embedded into a background of gold gold colliding at 0 – 4fm.	21
5.5	Transverse isolation energy of direct photon clusters calculated using cone sizes of 0.2, 0.3, 0.4. Photon jet pairs generated in Pythia embedded into a background of gold gold colliding at 4 – 8fm.	21
5.6	Comparison of reconstructed isolation energy (found by the detector) and the true isolation energy (found using known simulation particle information).	22
6.1	MultiLayer Perceptron with 4 total layers, in this case the user would have specified 6 input variables, 2 hidden layers, and 4 nodes per hidden layer.	24
6.2	Boosted Decision Tree showing the cut values. The output of this tree would be weighted among a forest of random BDTs [2].	25
6.3	Example for kNN for 3 input variables (3 dimensional space) [2]. 3 plots show projections of the planes, the circles show the search area for the nearest neighbor to the test event, this area is adjusted if k neighbors are not found.	25
7.1	Diagram showing how the PHENIX photon/electron identifying algorithm extracts the variables to cut on.	27
7.2	The energy distribution between towers of clusters from a direct photon, decay photon, and a charged hadron to show how different charged hadrons appear.	28
7.3	Roc Curves quantifying the effectiveness of several machine learning algorithms. Signal efficiency is what fraction of a sample of events is direct photons, background rejection is what fraction of the background signal is being removed from a sample of events. The 5 GeV $< E_T < 40$ GeV for signal and background clusters	29
7.4	Diagram showing π^0 decays to 2 photons when the π^0 has approximately 5GeV (a), 15GeV (b), 30GeV (c).	30
7.5	Roc Curves quantifying the effectiveness of several machine learning algorithms for clusters with 5 – 15GeV of energy	31
7.6	Roc Curves quantifying the effectiveness of several machine learning algorithms for clusters with 15 – 25GeV of energy	31
7.7	Roc Curves quantifying the effectiveness of several machine learning algorithms for clusters with 25 – 40GeV of energy	31

Acknowledgements

The research and algorithms developed in this study would not have been possible without the assistance of many supportive individuals. I would like to extend my gratitude first to my research advisor Dennis Perepelitsa for his guidance over the course of my undergraduate research studies, and for giving me the opportunity to contribute to the research efforts of the Heavy Ions physics community. Additionally I would like to thank all the members of the Heavy Ions Lab at the University of Colorado at Boulder for their assistance and the knowledge they imparted over the course of my time in the lab. I would like to acknowledge Francesco Vassalli, a member the CU Heavy Ions lab, for his collaboration and coding support over the course of this thesis research. I would also like to extend my gratitude to John Cumalat, my honors counsel representative, and Dirk Grunwald, my out of department committee member, for taking the time to read my thesis in depth as well as for attending my thesis defense. I would like to thank the members of the sPHENIX Collaboration for helping me contribute to sPHENIX Simulation efforts; specifically I would like to thank Jin Huang at Brookhaven National Laboratory for his help on integrating the cluster isolation algorithm into the sPHENIX simulation software. Finally I would like to extend my deepest gratitude to my parents Todd and Gayle Smith as well as the rest of my family. Without their love and support I would not have been able to complete this research, let alone attend the University of Colorado at Boulder to attain my degree in Engineering Physics.

Abstract

The sPHENIX experiment at Brookhaven National Laboratory's Relativistic Heavy Ion Collider (RHIC) will be a new detector with the purpose of studying the strongly interacting matter created in high energy Au+Au collisions. The high energy density created in Au+Au collisions produces a medium of disassociated quarks and gluons called Quark Gluon Plasma (QGP), which is the same substance which existed in the first microsecond of the universe. As a result of its incredibly strong interactions QGP is difficult to study; however, using the phenomena of jet quenching with photon jet pairs it is possible to gain insight into its nature. In order to use photon jet pairs to study QGP, the direct photon that is opposite the jet must first be correctly identified. Several algorithms are developed that use the sPHENIX calorimeters to reconstruct, identify, and isolate high energy direct photons that are created during the hard subprocess in high multiplicity Au+Au events. In this study events simulated by Pythia8 containing a photon jet pair are embedded into central Au+Au HIJING events and a full GEANT4 simulation of the sPHENIX detector is performed. Using reconstructed data from the simulated detector an algorithm is developed to determine how isolated an electromagnetic cluster is; direct photons tend to be more isolated than background sources. An algorithm is also developed to determine the probability an electromagnetic cluster resulted from a direct photon rather than a hadron or photon producing decay. This algorithm takes tower information from clusters in the electromagnetic calorimeter, extracts information related to the shape and energy of the showers, and passes those variables to a trained machine learning algorithm. The machine learning algorithm then outputs the clusters likelihood of being a direct photon. Using the cluster isolation and identification it is possible to identify direct photons.

Chapter 1

High Energy Collisions

1.1 Proton Proton Collisions

Before discussing heavy ion collisions, it is helpful to first gain an intuition for a much simpler system: the proton proton collision. But, first we must ask: why collide protons? When a proton is accelerated to large velocities it gains a large amount of kinetic energy and when it is collided with another proton they interact and produce a shower of particles. The collisions contain so much energy that, according to the relation $E = mc^2$, many hadrons can be produced during the interaction. The resulting spray of hadrons contains hadrons commonly seen today as well as exotic hadrons that only existed moments after the Big Bang. By observing the resulting spray of particles from these collisions it is possible to glean information about not only the universe today, but also the universe 13.8 billion years ago.

Protons are collided in experiments using two primary methods: in the first method 2 beams of protons, more like diffuse clouds, are shot towards each other in a collider and a few proton proton interactions may occur. In the second method a beam of protons is shot at a “target” of protons. For this study the method of shooting 2 beams of protons at each other is used. Before the protons are collided each of them is accelerated to relativistic speeds, near the speed of light, using electric fields, magnetic fields, or a combination of the two. After the protons have reached a kinetic energy desired by the experimenter their paths are crossed and they collide. The nature of this collision is dependent on the proximity of the particles as well as their energy.

1.2 Hard Subprocess

During low energy proton proton collisions the protons scatter elastically similar to when the cue ball collides with another in pool; however, as the energies of the colliding particles increase the collisions begin to probe the proton’s internal structure: quarks and gluons. Instead of the protons scattering off each other, a quark or gluon from each of the protons will interact through the strong force resulting in new particles with changed momentum and energy. This interaction is what is called the “hard subprocess”. The strong interactions that occur during the hard subprocess can be difficult to imagine so we introduce Feynman Diagrams as a useful tool to illustrate these interactions. Figure 1.1 shows an example of a

proton proton collision where a quark from each of the protons interacts strongly via a gluon then emerges with a modified energy and momentum.

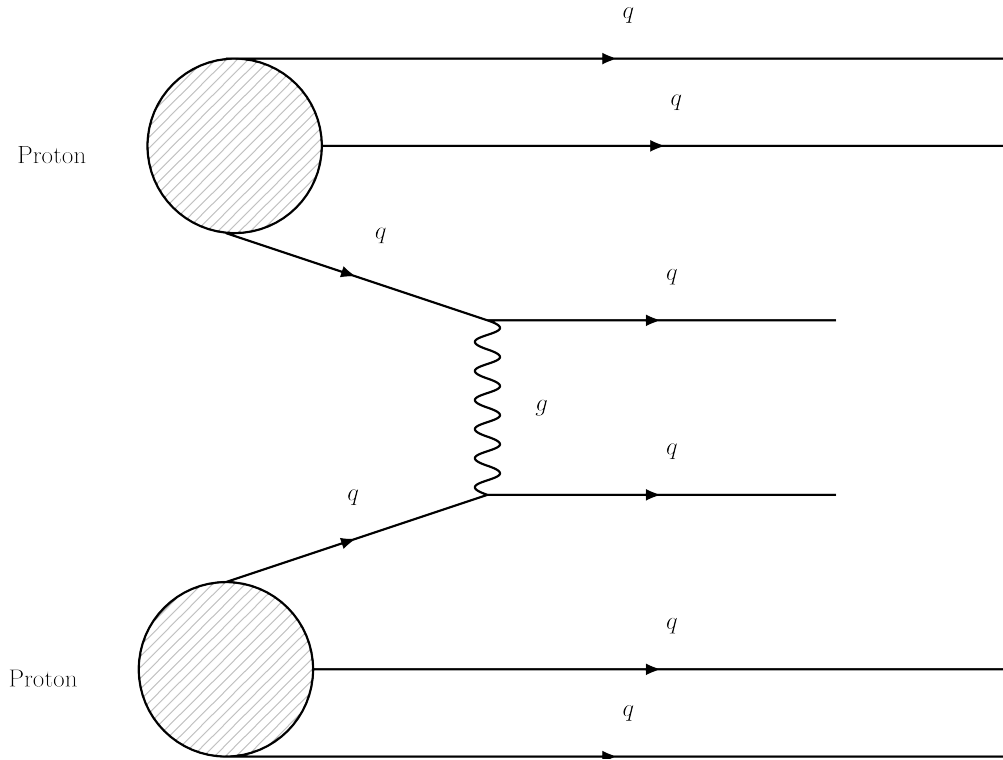


Figure 1.1: Example of two quarks interacting strongly during a proton proton collision.

Feynman diagrams are often shown only with the strong interactions and do not show the origin of the quarks or gluons. Figure 1.2 shows 3 examples of the strong interactions that can occur during the hard subprocess in proton proton collision.

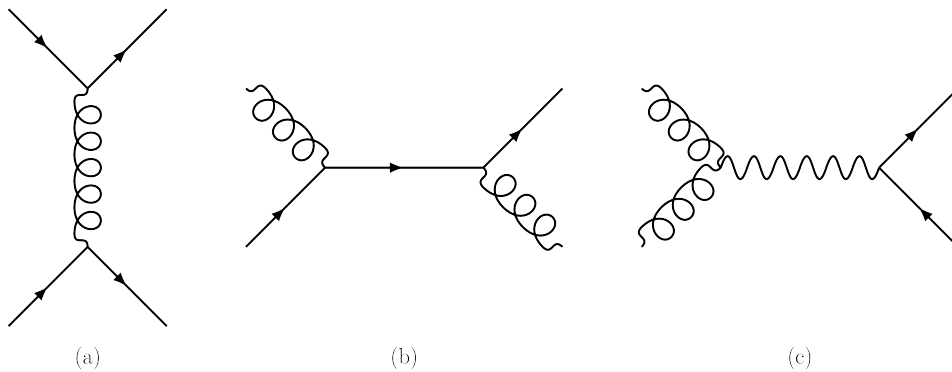


Figure 1.2: Two quarks interacting and producing two quarks (a), a quark and a gluon interacting to produce a quark and a gluon (b), two gluons interacting and producing a quark and antiquark pair (c).

1.3 Jets

Each of the quarks or gluons that interact have a portion of the proton's overall momentum when they collide. After the collisions conservation of momentum must be maintained so the particles may have different momentum along the direction they were originally travelling; however, in the plane perpendicular to their original direction, known as the transverse plane, the momentum of the two particles exiting the hard interaction will be exactly equal in magnitude and opposite in direction. Figure 1.3 illustrates this, showing two quarks during the hard interaction and those same quarks exiting the hard interaction back to back with equal transverse momentum.

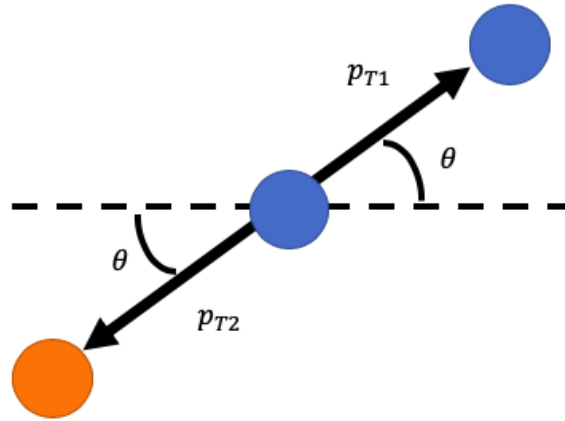


Figure 1.3: Quark (Blue) travelling into page, quark (orange) travelling out of page.

Almost immediately after the quarks or gluons exit the collision the quarks emit gluons and the gluons emit quark anti-quark pairs or gluons. This process continues as quarks and gluons are emitted from the emitted quarks and gluons creating a cascading effect that results in greater numbers of gluons and quarks (Figure 1.4). The travelling quarks and gluons will then hadronize, grouping together into colorless particles such as meson and baryons. The resulting mix of hadrons will then decay into other hadrons continuing this cascading effect of creating more particles.

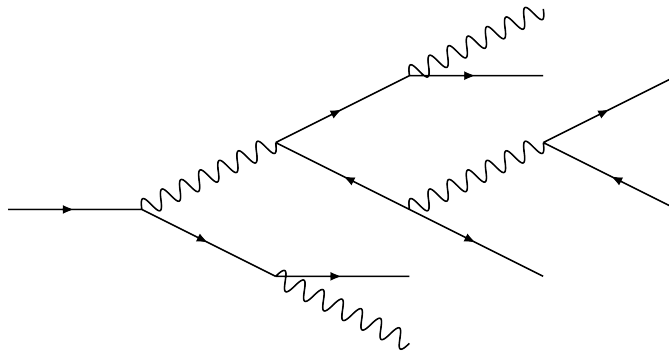


Figure 1.4: A quark exiting the hard subprocess emitting a gluon, and creating a shower of quarks and gluons that will become a jet.

A particle during this evolution will emit or decay into a particle that travels at a relatively small angle from the direction that the first particle is travelling. Think of throwing a baseball out the window of a moving car, the baseball will still be traveling in a similar direction to the car regardless of the original direction it was thrown. This results in a grouping of particles traveling in a cone centered around the original direction of the quark or gluon that exited the hard subprocess. The resulting group of particles all traveling within a given cone is what is called a jet.

Jets are a useful tool that can be used by physicists to study the original quarks or gluons that created them, they can be used to study the branching process that occurs while jets are being made, and they can be used as a probe to study the medium that the jets travel through. A jet will most often appear with either another jet across from it in the transverse plane to balance its momentum; although, more exotic situations involving high energy particles, 2 jets, or more jets do occur.

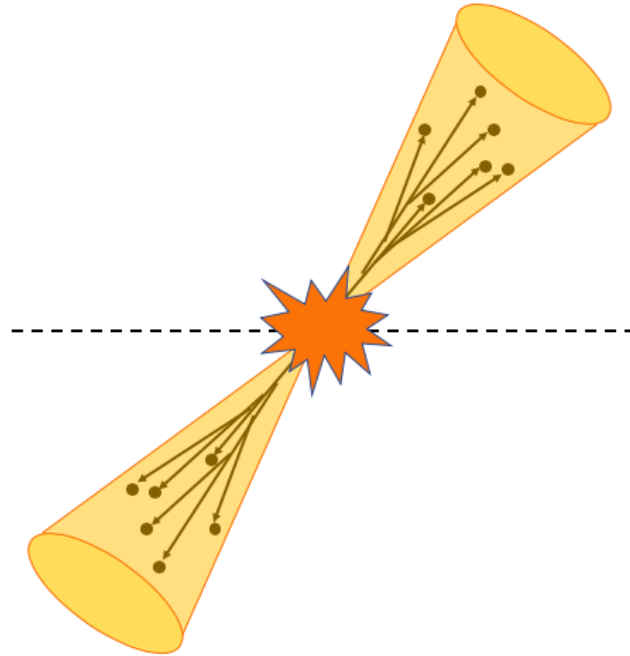


Figure 1.5: Two jets exiting a proton proton collision, known as a dijet event

1.4 Photon Jet Pairs

The examples of possible hard interactions that can occur shown in Figure 1.2 only showed quarks and gluons exiting the hard subprocess; however, this is not always the case. It is possible for a photon to exit the hard subprocess opposite a quark or gluon as shown below in Figure 1.6.

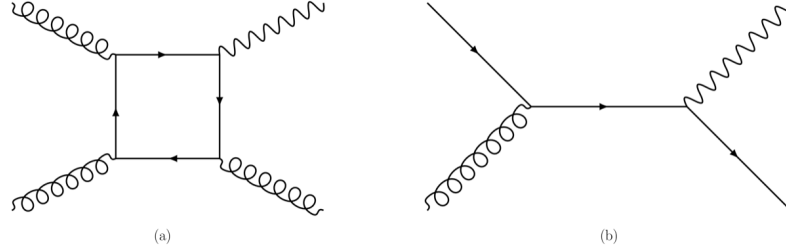


Figure 1.6: Two gluons interacting to form a photon and a gluon (a), a quark and gluon interacting to form a quark and photon (b).

Photons, unlike quark or gluons, will not emit other particles unless they interact with matter. This means that a photon will travel with its original direction and momentum until it encounters, in most cases, a detector designed to measure photon's energy and direction. Recall that the particles exiting the hard interaction will have equal and opposite transverse momentum. The photon and quark or gluon will be initially back to back, which presents the unique opportunity of being able to study the jet opposite the photon before fragmentation, hadronization, or before the jet particles interact with any medium through which they are passing.

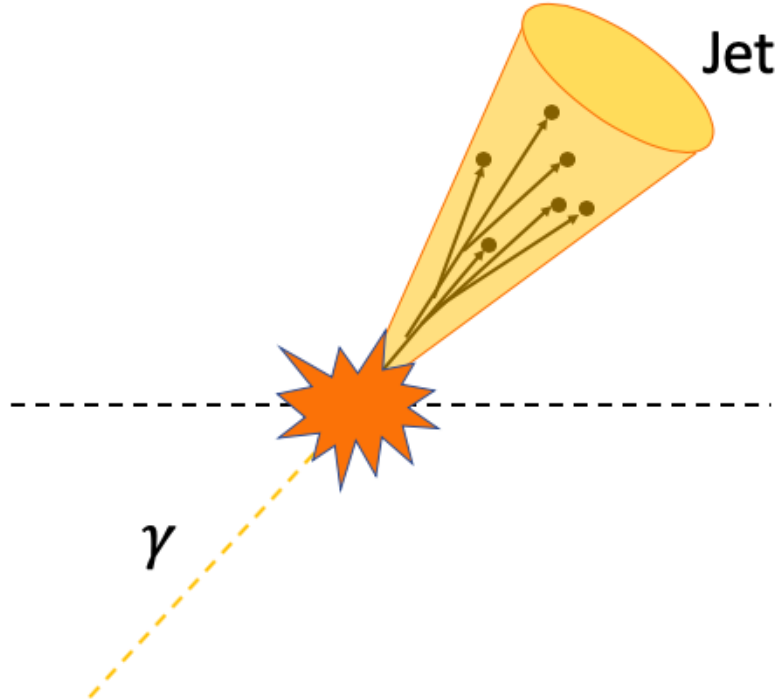


Figure 1.7: A photon jet pair from a proton proton collision.

Chapter 2

Heavy Ion Collisions

2.1 Heavy Ion Collisions

Now that the proton proton collision system has been introduced, the more complicated heavy ion collision system can be more easily discussed. But, just as with proton proton collisions, we first must ask: why collide heavy ions? Moments after the birth of the universe, it rapidly expanded from an infinitely small, infinitely hot point as shown in Figure 2.1.

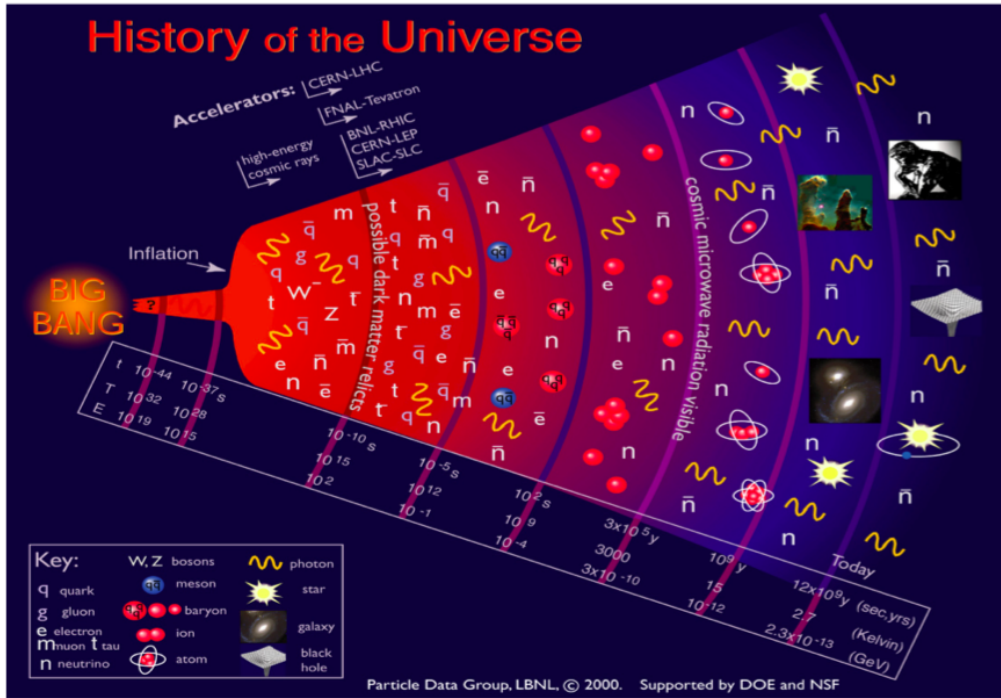


Figure 2.1: The history of the universe as told from a particle point of view.

Between 10^{-37} and 10^{-5} seconds after the Big Bang (Figure 2.1) the temperature and density of baryonic matter were such that hadrons were unable to form and instead the universe was a continuous medium of quarks and gluons called Quark Gluon Plasma (GQP).

By accelerating fully ionized nuclei with a high atomic number, $Z = 82$ for Pb, $Z = 79$ for Au, to energies on the order of 100 GeV and colliding them, a high

density and temperature environment is created such that Quark Gluon Plasma comes into existence. Studying the resulting particles and jets that travel through this medium allow us to learn about QGP and in doing so glean information about the nature of the universe and its behavior shortly after its creation.

2.2 Quark Gluon Plasma

Since its discovery in Brookhaven National Laboratory's Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC) at CERN [3], Quark Gluon Plasma has been consistently generated and studied through gold gold collisions at RHIC and lead lead collisions at the LHC. As shown in Figure 2.2, QGP spans a large range of temperatures and baryon chemical potentials. Different experiments probe different regions of this QCD phase diagram and allow a glimpse into the nature of this enigmatic state of matter.

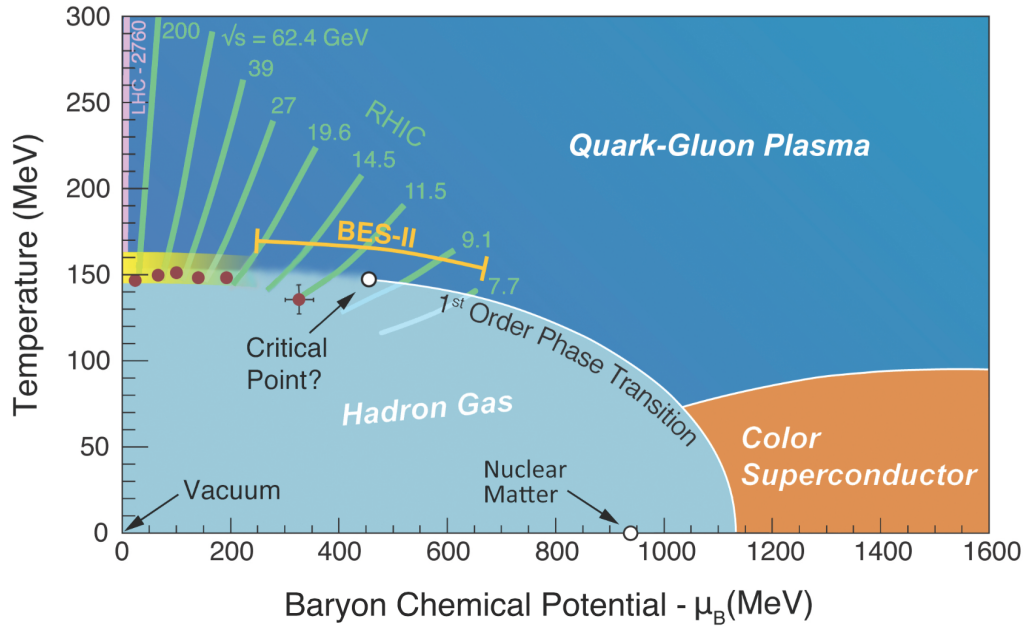


Figure 2.2: QCD Phase diagram showing states of matter and where different experiments probe the diagram.

As an example studies of QGP have yielded that it doesn't exactly behave as a gas of totally dissociated quark and gluons; it behaves as if the quarks are strongly coupled to their neighboring quarks by gluons but are free to move around the medium. This is analogous to how ions and electrons behave in a plasma and the reason that QGP is referred to as a plasma of quarks and gluons [3]. However, there are many questions left to be answered. How does QGP form and then freeze out in $1 \frac{fm}{c}$? Does QGP significantly change as it has more quarks or gluons, or as more net quarks over antiquarks? How does QGP change with increasing temperature and density?

2.3 Jet Quenching

One of the most important tools for studying the properties of quark gluon plasma in heavy ion collisions is jets. When jets pass through QGP the quark or gluon that starts the jet, as well as the quarks and gluons and hadrons that will later be in its jets, interact strongly with the medium. As a result of interactions with the QGP the initial momentum and direction of the jet is modified; this modification is known as jet quenching.

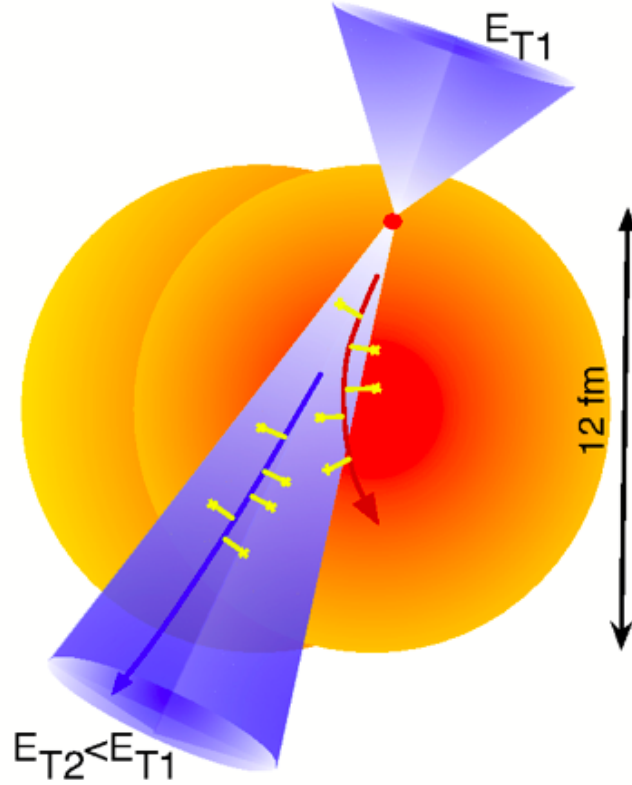


Figure 2.3: Jets being quenched differently depending on the distance they travel through quark gluon plasma.

Recall in Section 1.4 that when there is a photon jet pair the photon will have equal and opposite transverse momentum and direction to the jet it forms with. In the case when QGP is present the jet will be quenched, but the photon which does not interact strongly will sail right through the medium with little to no modification to its direction and momentum. If the photons final transverse momentum is known and the jets final transverse momentum is known then it is possible to study the effects that QGP has on strongly interacting matter.

2.4 Background

Heavy ion collisions provide an opportunity to learn more about many fields of physics, but do not come without their challenges. The sheer number of nucleons participating in the collision as it graduates from proton proton collisions (2 nucleons) to gold gold collisions (394 nucleons) makes the environment crowded, and

that is before considering the energies involved. The result is a quadratically larger number of particles that make up what is called a high multiplicity background; however, the background is not as simple as a certain number of overlapping proton proton events. All of the particles interact and behave as a collective, giving the event an overall "shape". In the most central heavy ion collisions there is typically 1000s of particles per unit rapidity (angle along the beam direction). In order to observe phenomena such as jet quenching it becomes necessary to do things like subtract an energy pedestal, or even recalibrate the detectors to deal with the larger amount of energy.

Chapter 3

sPHENIX

3.1 Useful Values

Before discussing details about detectors it will be helpful to go over common variables and units used when discussing detectors and collisions within them.

Pseudorapidity ($\eta = -\ln[\tan(\frac{\theta}{2})]$) references an angle with respect to the direction the beam is traveling where 0 is exactly perpendicular to the beam direction and ∞ is along the beam direction as shown in Figure 3.1.

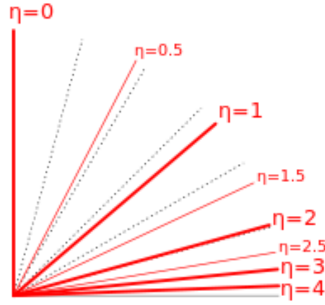


Figure 3.1: Diagram showing different values of pseudorapidity.

Azimuthal angle (ϕ) refers to the angle in the plane perpendicular to the beam direction. ϕ goes from $0-2\pi$, though due to the symmetry of the sPHENIX detector there is not necessarily a place set to 0 and analysis can be done in different ranges of ϕ if convenient.

Transverse momentum (p_T) and transverse energy (E_T) refer to the momentum or energy of a particle in the plane that is perpendicular to the beam direction. This is commonly used because momentum must be conserved so things will be back to back in the transverse plan, as well as the large amount of energy along the beam direction make dealing with values in the transverse plane easier to use in calculation.

MeV/GeV/TeV stand for mega electron Volts, giga electron volts, and tera electron volts accordingly. Often these values will be used in reference to mass, momentum, and energy despite those values not having the same units. For example when referring to MeV , what is really meant is that if it is mass it has units of $\frac{MeV}{c^2}$, momentum is $\frac{MeV}{c}$, and energy is MeV . This convention comes from setting $\hbar$ and c equal to 1

3.2 RHIC and PHENIX

This study pertains specifically to experiments conducted at the sPHENIX detector that will be built at the Relativistic Heavy Ion Collider (RHIC). The Relativistic Heavy Ion Collider is a particle accelerator at Brookhaven National Laboratory in New York designed to collide protons and ionized gold nuclei at center of mass energies up to 200 GeV per nucleon. Collisions at this energy allow physicists around the world to study nuclear physics, particle physics, astrophysics, condensed matter, and cosmology.

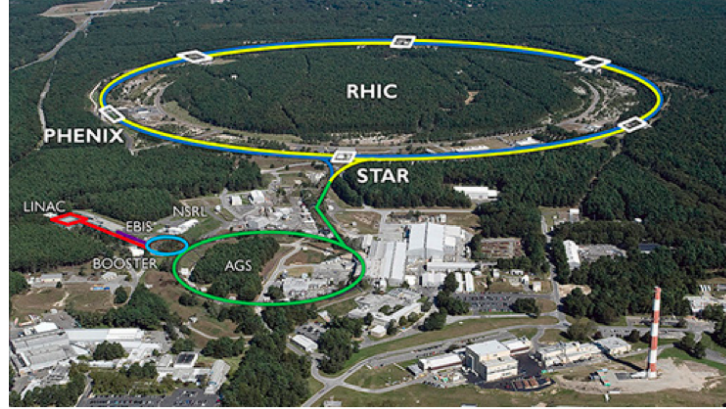


Figure 3.2: Aerial Photograph of the Relativistic Heavy Ion Collider

At RHIC, electric and magnetic fields accelerate and guide two beams of particles travelling opposite directions around a 2.4 mile circular 'track' [4]. Each of these beams is traveling in its own beampipe and only cross at 6 locations. One of these locations is where the PHENIX detector currently lies. So named for its distinctive birdlike cross section, the 4000 ton PHENIX detector was designed to study the quark gluon plasma produced in collisions by RHIC by observing high- p_T hadrons, leptons, and photons that exit the collision [4].

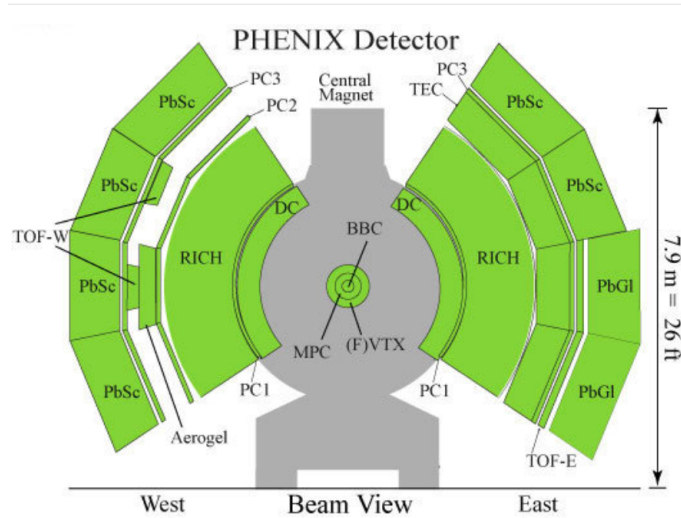


Figure 3.3: Cross section view along the beam line of PHENIX and the detector that compose each of its arms [1].

Detectors do not last forever; as accelerator technology pushes past what a detector can handle and as detector technologies themselves advance, older detectors slowly become less useful. For this reason the sPHENIX detector is being developed to take the place of the PHENIX detector at RHIC.

3.3 sPHENIX

sPHENIX is a future detector designed to probe quark gluon plasma primarily through the study of jets. It covers $0 - 2\pi$ in ϕ , ± 1.2 in pseudorapidity η , and reads out data at a rate of 15kHz [5]. sPHENIX is composed of 4 different detectors that are arranged as layers of a cylinder around the beam pipe as shown in Figure 3.4.

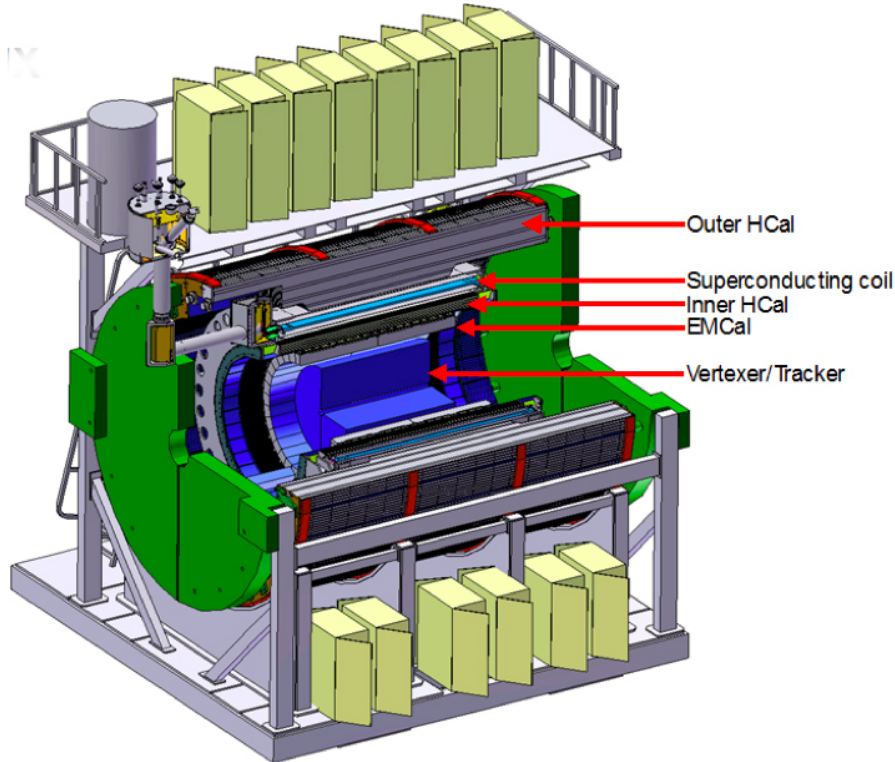


Figure 3.4: CAD rendering of sPHENIX and the detectors that compose each of its layers [1].

Just around the beam pipe is the tracker. When charged particles exit collisions in the detector they pass through the tracker and leave small ionized hits in their wake. The tracker is designed to detect these hits, reconstruct the particle's track, and determine the momentum and identity of particle. The trackers effectiveness is highly dependent on the magnet which will be discussed shortly.

Around the tracker lies the electromagnetic calorimeter (EMCal). Designed to absorb and detect the energy of electromagnetically interacting particles the sPHENIX EMCal is made up of 24576 towers with a single layer of 96 towers along the η direction and 256 towers along the ϕ direction. Each tower measures 0.025 in η and 0.025 in ϕ and is made up of 677 scintillating fibers embedded in tungsten powder and epoxy [6]. When electromagnetically interacting particles come into contact with the EMCal they interact with electrons in the material, producing photons

which produce electrons and positrons, which produce more photons in an exponential fashion that acts like an avalanche of electrons and photons rushing through the detector material. This phenomena is known as electromagnetic shower. The scintillating fibers throughout the material then detect the shower and through some computation the energy deposited in the detector can be determined to a certain resolution. Electrons, positrons, and photons deposit nearly all of their energy in the EMCal while charged hadrons deposit some energy but pass still through.

Outside the electromagnetic calorimeter is the inner hadronic calorimeter (inner HCal). Designed to absorb and detect the energy of strong interacting particles, this inner hadronic calorimeter is made up of steel plates interleaved with scintillator tiles that measure 0.1 in η and 0.1 in ϕ . When charged and neutral hadrons pass through the inner HCal they interact strongly with the steel plates in the detector towers producing a hadronic shower. As this study focuses on the EMCal the exact process of this shower won't be focused on. The scintillating tiles then sample the energy deposited in the tower and through computation the actual energy deposited can be determined.

Around the inner hadronic calorimeter is the reason that there is an inner and outer hadronic calorimeter: the magnet. sPHENIX will be using the 1.5 Tesla [5] superconducting magnet that the Babar experiment used previously. Because the magnet has been re-purposed its size cannot be changed and as a result the hadronic calorimeter needs to be built in 2 parts around the magnet. A very strong magnet is important component of a detector because without it the tracker would be useless. The magnetic field generated causes the particle travelling through the tracker to bend and with this bending the identity and momentum of the particles can be determined; however, if the magnetic field is not strong enough the more energetic particles will appear to go straight regardless of their charge and the tracker is less helpful in measuring their momentum.

Outside of the superconducting magnet lies the final detector: the out hadronic calorimeter (outer HCal). Is it made up of the same steel interleaved with scintillating tiles as the inner HCal but the towers are slightly longer to make sure as many particles as possible are absorbed and their energy is well measured. Additionally the towers are aligned at an opposite angle to the inner HCal to confirm a particle is absorbed even if it bends past the steel plates in some of the towers.

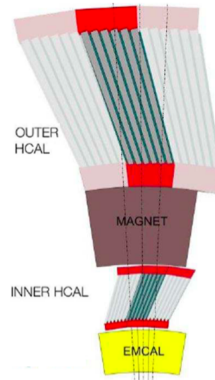


Figure 3.5: Slice of the sPHENIX calorimeters and magnet showing opposite alignment of inner and outer HCal towers.

3.4 Clusters

As particles shower in the electromagnetic and hadronic calorimeters, the particles that make up the shower will spread out similar to how jets spread into a cone. The spread of particles results in a grouping of towers with higher energy at the middle and lower energies further from where the particle initially impacted. Grouping together these towers is important because they resulted from the same particle impact, and the way they are grouped together is by combining them into a cluster. Ideally, a cluster is a group of towers whose energy content results from the same particle. Their attributes include the number of towers in the cluster, the total energy, the position in the detector, and the distribution of energy between towers in the cluster. Different particles at different energies will have a different cluster shape determined by this distribution. These cluster variables will be used for isolation and identification in this study.

Chapter 4

Simulation

4.1 Generating Events

To generate events for this study Pythia 8.1 [7] was used, where each collision is considered an event. Proton Proton collisions were simulated at $E_{CM} = 200$ GeV with a minimum particle transverse momentum of 5 GeV. All HardQCD processes were allowed in addition to all promptPhoton processes. HardQCD is generally used for dijet events, events with 2 jets exiting the scattering opposite each other. PromptPhoton is generally used for photon+jet processes. A random seed was set such that no artificial patterns would emerge in the simulated events. Events were then simulated and only ones which had a minimum transverse energy exchange during the hard scattering of 5 GeV were kept. This corresponds to the final state p_T of the photons and jets exiting the hard scattering. The event production process was repeated until a large number of photon jet pair events were accrued.

4.2 Gold Gold Background

Simulating gold gold collisions in large numbers would take a very long time to do with the resources available; however, the sPHENIX collaboration has generated pre-simulated backgrounds that result from gold gold collisions. These large backgrounds are generated using HIJING [8]: a perturbative-QCD based Monte Carlo generator for producing proton proton, proton nucleus, and nucleus nucleus events. Having pre-generated gold gold backgrounds allows the embedding of the photon jet pair events into the gold gold background. This allows one to analyze the event as if it were truly a gold gold collision.

4.3 Detector Simulation

Now that proton proton and gold gold events have been generated they need to be simulated in the detector. This is accomplished through a combination of the sPHENIX Software Framework and GEANT4 [9]. The sPHENIX Software Framework allows the translation of truth information from Pythia8 and HIJING simulations into reconstructed detector information (Figure 4.1). The geometry of the detector, the materials it is made of, the parts of the detector present, and the

truth information from the event files are all set with the software framework. Using GEANT4 the materials of the detector and how they interact with particles are then simulated. The result is an output of reconstructed tracks, calorimeter energies, and truth information with which to do analysis, efficiency studies, or to develop direct photon isolation and identification algorithms.

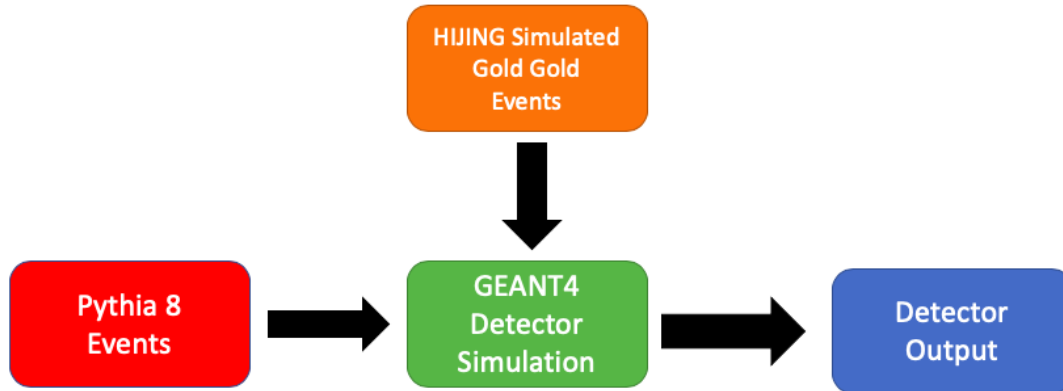


Figure 4.1: The flow of computer simulation. Pythia8 events are produced and along with already simulated gold gold event are put into a full detector simulation which then outputs truth and reconstructed information about the event.

Chapter 5

Cluster Isolation

5.1 Background Signals

Now that a number of events have been simulated that contain direct photons the question arises of how to identify this very specific kind of photon. The first thing that needs to be known before direct photons can be separated from background signals is what are the background signals?

The first and easiest background signal to identify and account for is thermal photons. QGP commonly exceeds temperatures of 200 MeV, around 2.3×10^{12} Kelvin, at RHIC (Figure 2.2) and just like the Sun, or any other hot substance, it radiates some of its energy as light, the photons that make up this light are referred to as thermal photons. Through studies [10] it has been found that QGP at 200 MeV have emission rates on order of 10^{-9} and below for 3 GeV photons, which is fortunate as the direct photons this study is concerned with identifying are 5 GeV and greater. This is because identifying any kind of jet that is below 5 GeV is very difficult in the high multiplicity background generated in gold gold collisions. Due to the low number of thermal photons at energies at and above 5 GeV, by instituting an energy cut for what photons to consider as direct it is possible to eliminate the nearly all of background from thermal photons without reducing the usable direct photon statistics.

A second source of background signal is electrons which shower similarly in the electromagnetic calorimeter to photons. While similar, this signal is reasonably easy to account for because electrons have a bent trajectory in the sPHENIX tracker due to the magnetic field and photons do not because they are uncharged. As a result, electrons will not be heavily considered as a background signal in this study.

The third and more difficult background signal to identify and account for are decay photons. Just fermi, 10^{-25} seconds, after the QGP forms, it cools and the quarks and gluons hadronize to forms particles such as mesons and baryons. Many of the resulting mesons and baryons decay into high energy photons which are often in the energy range of direct photons. Thus, arises the questions of how to differentiate between decay photons and direct photons because they both appear as high energy photons in the sPHENIX electromagnetic calorimeter. One method to try and differentiate these two kinds of photons is with cluster isolation.

Additionally there is a background signal resulting from charged hadrons that shower in the electromagnetic calorimeter but do not deposit all of their energy in it. This background signal can be removed similarly to electrons using the tracker,

also it can be removed using cluster isolation similar to decay photons.

5.2 Goals

Cluster isolation serves as a metric to test how many particles are around a cluster in the calorimeter, which is useful in identifying direct photons because they do not shower or emit many particles until they collide with the detector. As a simple example Figure 5.1 shows the extent to which a direct photon is isolated in a proton proton collision. In the figure the height of the colored bars represents the energy deposited in the calorimeter towers at a given angle ϕ . The direct photon can be seen in the EMCal on the left while the jet opposite it can be seen in all calorimeters on the right. Note the minimal energy deposited in towers around the direct photons.

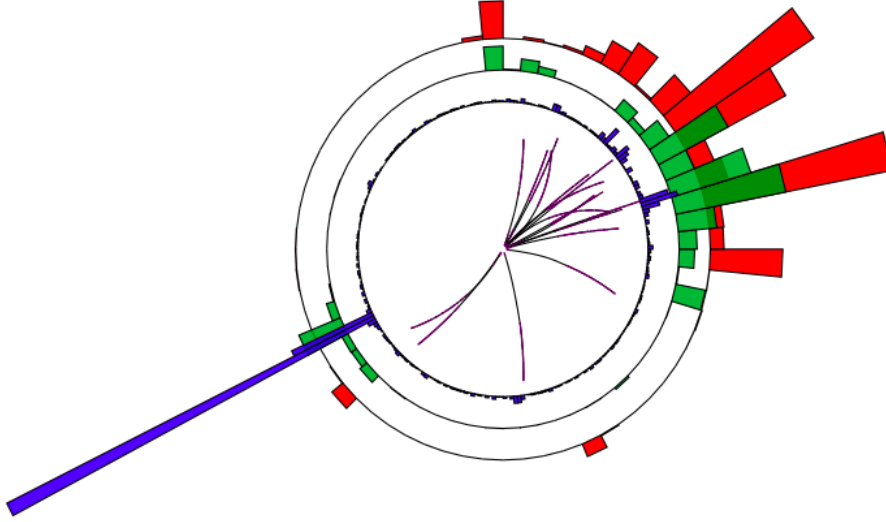


Figure 5.1: A side view of the sPHENIX tracker hits (purple dots), trajectories in the tracker (black), electromagnetic calorimeter (blue), inner hadronic calorimeter (green), and outer hadronic calorimeter (red).

In this study an algorithm is developed with the goal of calculating isolation energy of clusters in the electromagnetic calorimeter at sPHENIX in proton proton and gold gold collisions using only the electromagnetic and hadronic calorimeters. The isolation energy is calculated by summing all the transverse energy of the calorimeter towers inside a cone of a given size around the center of the cluster and then subtracting the transverse energy of the cluster itself. The equation used to calculate isolation energy is shown in equation 5.1.

$$E_T^{Isolation} = \sum_i^{R_{Tower} < R_{Cone}} E_T^{Tower,i} - E_T^{Cluster} \quad (5.1)$$

$$i = (EMCal, InnerHCal, OuterHCal)$$

5.3 Testing

The algorithm to determine isolation energy was first developed using proton proton collisions to test it due to the easier background and shorter simulations times. As a result of how Pythia simulates events several different ranges of photon energy needed to be tested separately. Figure 5.2 shows the results of testing the algorithm with direct photons and decay photons.

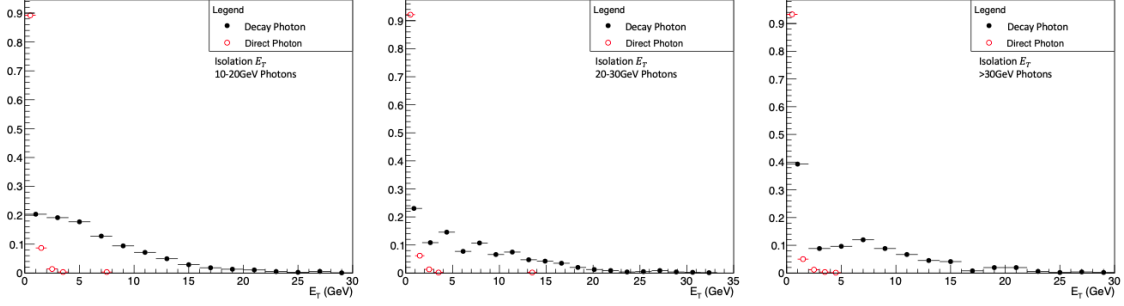


Figure 5.2: Transverse isolation energy of direct photon clusters and decay photon clusters using a cone size of $R = 0.3$.

In this simulation π^0 mesons were used as an example of decay photons because a π^0 decays to 2 photons 98.832% of the time after its mean lifetime of $8.30 \times 10^{-17} s$ [11]. As expected the majority of direct photons have an isolation energy under 1 GeV while the distribution of decay photons isolation energy is on average peaked closer to 5 GeV. This shows that decay photons can be reasonably separated from direct photons using isolation energy in proton proton collisions.

After the concept of isolation energy was proven in proton proton collisions, it was then developed with the higher multiplicity environment of gold gold collisions. The first step in graduating from a proton proton background to a gold gold background is to account for the pedestal of energy that all the towers stand on. In gold gold collisions the high multiplicity background contributes additional energy to the calorimeter towers in addition to energy contributed by the jets and direct photons. It is important to note that the addition background contribution will not be consistent in the entire detector as a result of the events having a total "shape".

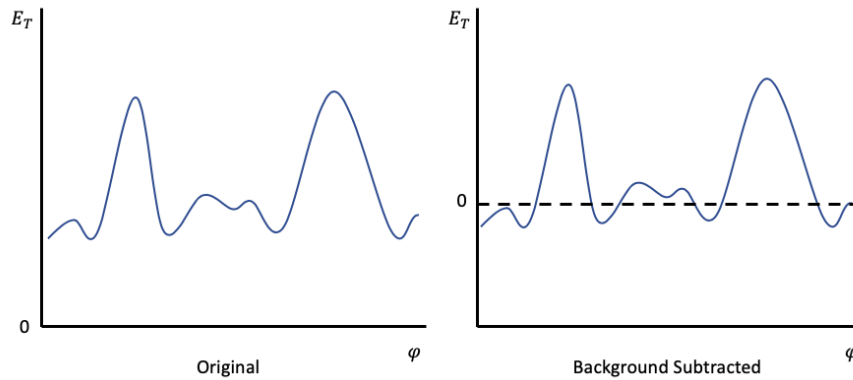


Figure 5.3: Cartoon showing in principle the background subtraction that is run on gold gold events.

Fortunately the sPHENIX software framework has an algorithm in place which can be run on events simulated in the detector which can calculate this pedestal of energy and subtract it from each of the towers in the calorimeters (Figure 5.3). Therefore to run the cluster isolation algorithm on gold gold events the background subtraction just needs to be called first. Figure 5.4 and Figure 5.5 show the cluster isolation algorithm being run with and without the background subtraction. Note that while the distributions become centered around 0, which is expected for direct photons, the distributions trail into negative isolation energies. This is a result of background subtraction being an average and is expected.

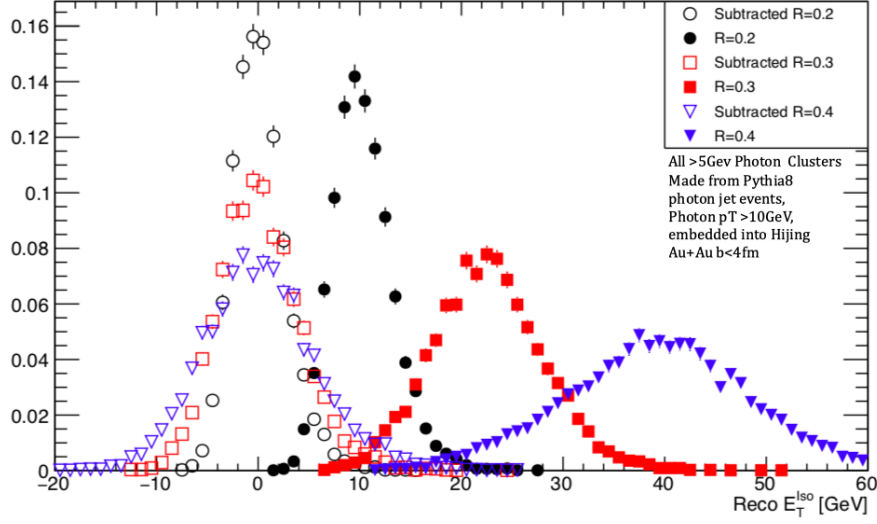


Figure 5.4: Transverse isolation energy of direct photon clusters calculated using cone sizes of 0.2, 0.3, 0.4. Photon jet pairs generated in Pythia embedded into a background of gold gold colliding at 0 – 4fm.

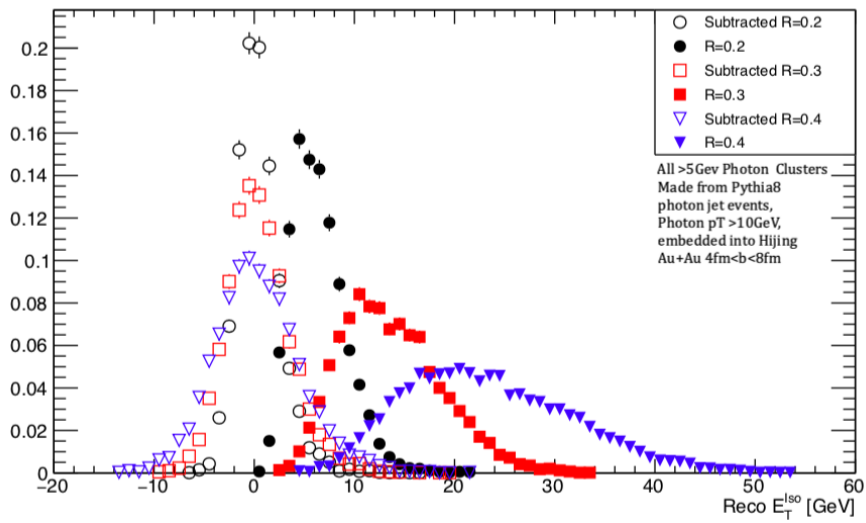


Figure 5.5: Transverse isolation energy of direct photon clusters calculated using cone sizes of 0.2, 0.3, 0.4. Photon jet pairs generated in Pythia embedded into a background of gold gold colliding at 4 – 8fm.

5.4 Moving forward

The cluster isolation algorithm has been proven to calculate cluster isolation energy in proton proton collisions and gold gold collisions for different cone sizes, without background subtraction, and with background subtraction. This algorithm is a helpful tool in identifying direct photons in the high multiplicity gold gold background, as well as a useful tool to understand the detector response to photons and photon backgrounds. For example it can be used to test the isolation energy found by the detector as a function of the actual isolation energy (Figure 5.6), which allows researchers to study the actual isolation energy given what the detector finds.

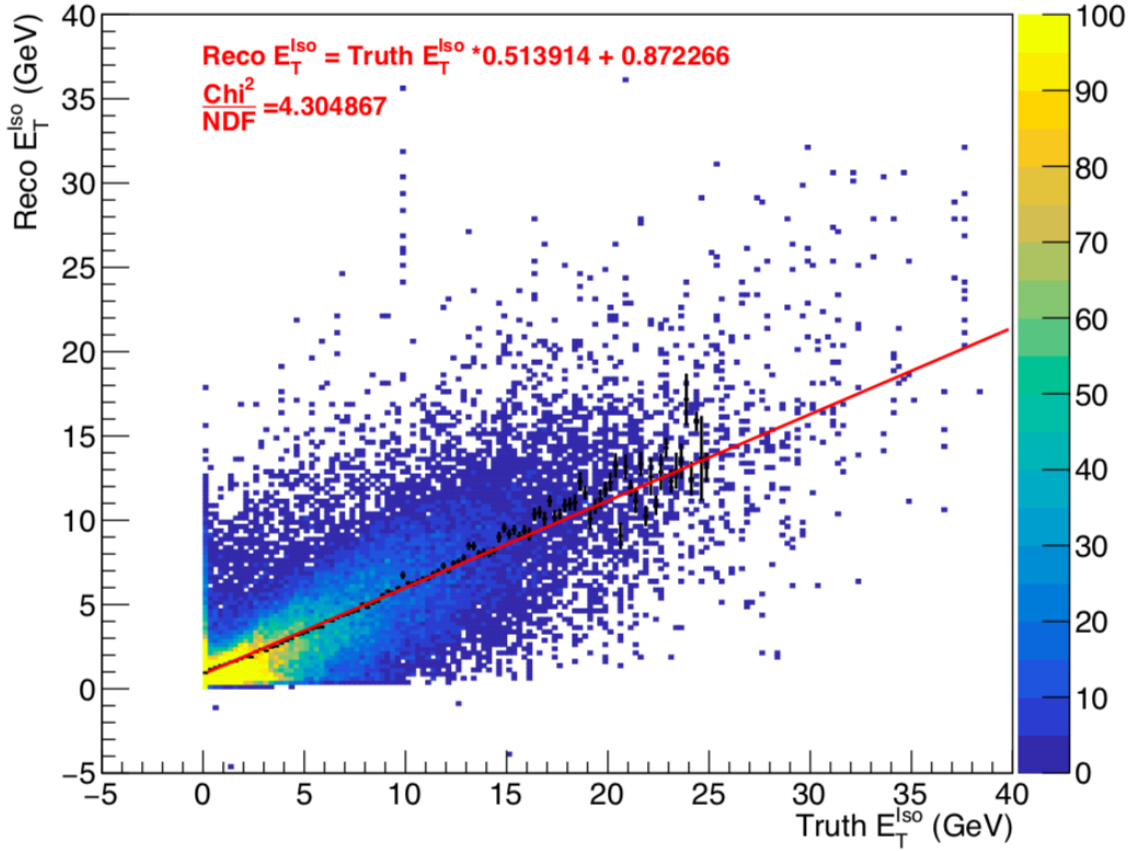


Figure 5.6: Comparison of reconstructed isolation energy (found by the detector) and the true isolation energy (found using known simulation particle information).

Cluster isolation, however, can not identify direct photons on its own. At other experiments they identify direct photons using a series of cuts on variables such as isolation energy and cluster shower shape (the shape of how the energy is distributed among calorimeter towers), determining the variables and fine tuning the cuts themselves. What if instead of the experimenter determining all the cuts themselves, they left it to a computer?

Chapter 6

Machine Learning

6.1 What is Machine Learning?

Machine learning algorithms are a class of algorithm that become systematically more accurate without being explicitly programmed to do so. The basic idea is that the algorithm receives data, makes output based statistical analysis of that data, then receives more data and updates the output accordingly. This iteration continues to whenever the computer scientist, or program, has decided the training of the machine learning algorithm is complete. The final result is an accurate algorithm that relies on patterns in data rather than explicit instructions, and if the explicit instructions for an algorithm are to match a pattern then machine learning is the natural choice in place of conventional algorithms.

6.2 Toolkit for MultiVariate Analysis (TMVA)

The Toolkit for MultiVariate Analysis provides an environment for training, testing, and applying a host of machine learning algorithms. This is an important tool for high energy and nuclear physics since data sets are getting larger and patterns are becoming smaller and less distinguishable [2]. TMVA provides an easy way to produce trained machine learning algorithms including [2]:

- Rectangular cut optimisation
- Projective likelihood estimation
- Multi-dimensional likelihood estimation
- Linear and nonlinear discriminant analysis
- Artificial neural networks
- Support vector machine
- Boosted/bagged decision trees
- Predictive learning via rule ensembles

Each of these varieties of machine learning algorithms comes with between 2 – 5 variations and ends up resulting in a large number of algorithms as possible candidates for making cuts on cluster variables. Only 4 of these algorithms will be discussed at greater length for reasons that will become apparent in the following chapter.

6.3 MLP

A MultiLayer Perceptron (MLP) is a kind of artificial neural network; an artificial network is a collection of neurons (nodes) that take several variables as an input. The state of those neurons is modified by the input, and serves as input for other neuron in the network until the final neuron's state is modified. The final neuron's state serves as the output for the artificial neural network. TMVA's MLP is a simple and controlled artificial neural network. The user specifies the number of nodes in a layer of the MLP and the number of layers present. If the user were to put more than the necessary number of nodes the algorithm would favor fewer nodes by making a certain number weighted much less [2].

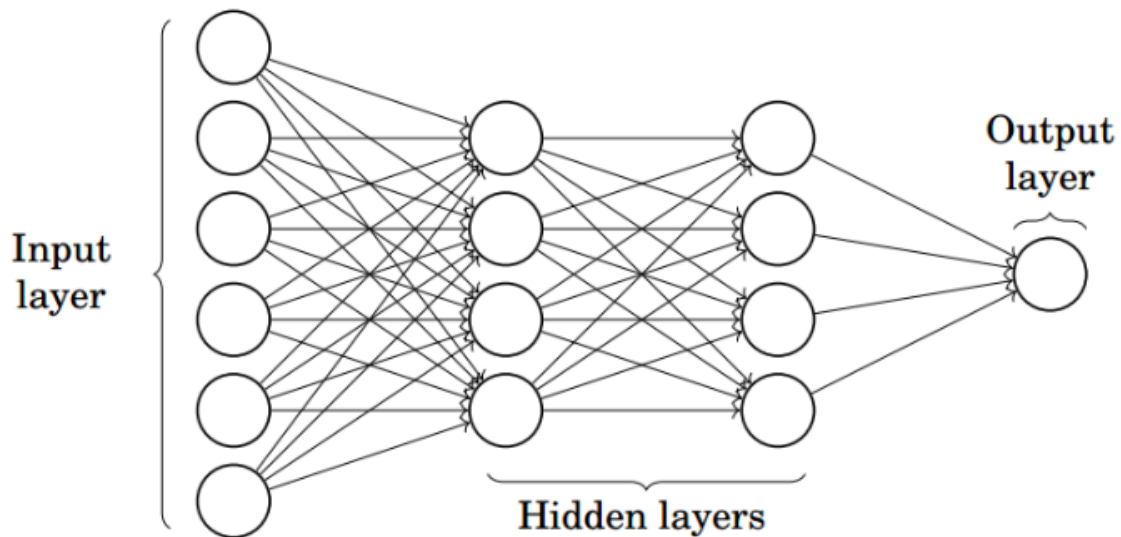


Figure 6.1: MultiLayer Perceptron with 4 total layers, in this case the user would have specified 6 input variables, 2 hidden layers, and 4 nodes per hidden layer.

6.4 BDT

A decision tree is a binary tree which makes cuts on a single variable at a time until the stop criteria is met and the tree either concludes the variables indicate signal or background. Alone binary trees are relatively weak, which is ameliorated by 'boosting' them. Boosted Decision Trees (BDT) are a machine algorithm technique where a random forest of binary trees is used which stabilizes response and enhances performance.

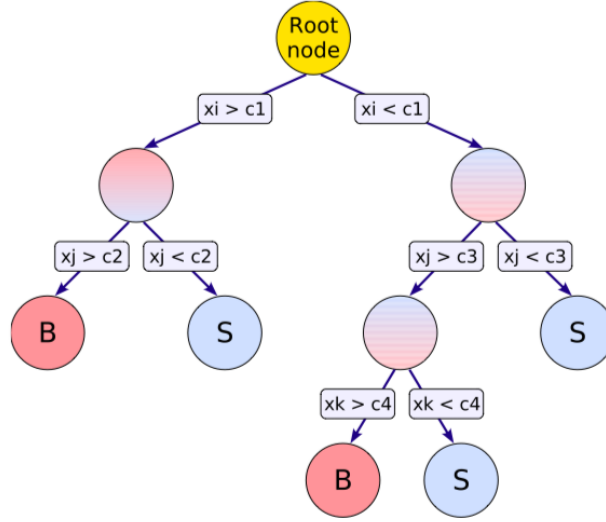


Figure 6.2: Boosted Decision Tree showing the cut values. The output of this tree would be weighted among a forest of random BDTs [2].

6.5 kNN

k-Nearest Neighbor compares an observed test event to reference events from a training set. The algorithm represents the number of variables as n dimensions, the data is represented as points in the n dimensional space, and the algorithm searches for k reference events that are the nearest neighbor of the test event. The algorithm then defines a volume in the n dimensional space to be used as a metric for signal [2]. This algorithm performs well when there is an irregular boundary between signal and background because it is intrinsically adaptive.

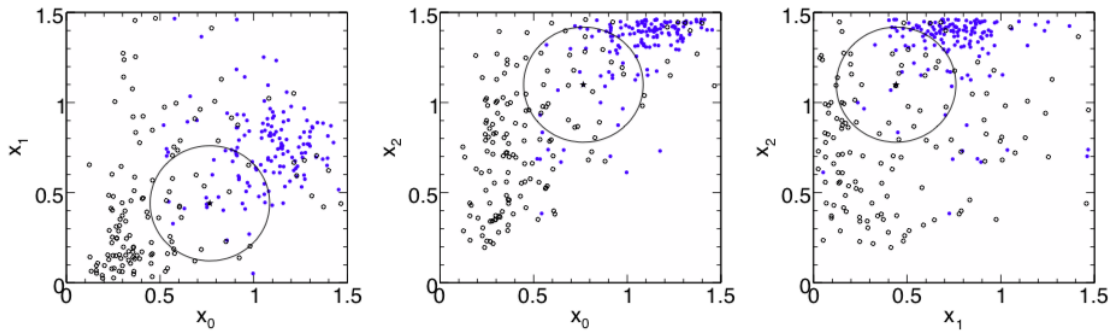


Figure 6.3: Example for kNN for 3 input variables (3 dimensional space) [2]. 3 plots show projections of the planes, the circles show the search area for the nearest neighbor to the test event, this area is adjusted if k neighbors are not found.

6.6 Likelihood

Likelihood is in many way a slightly simpler version of k-Nearest Neighbor. This method consists of building probability density functions (PDFs) for the signal and background input variables. For a given event the input variables are passed to the algorithm, the signal PDFs of all the input variables are multiplied and then normalized by the sum of signal and background likelihoods, and the result is a likelihood of being signal and not background. Note that this algorithm assumes all input variables are independent when multiplying the PDFs so it will not perform as well if the input variables are not orthogonal.

Chapter 7

Identifying Direct Photons

7.1 PHENIX Algorithm

As previously discussed, sPHENIX will be replacing the PHENIX experiment and it has inherited some of the PHENIX software. One piece of the PHENIX software in particular is an algorithm which determines the probability a cluster in the electromagnetic calorimeter is a single photon or electron, rather than a merged photon pair from a neutral hadron decay, or the start of a hadronic shower. While this is not exactly the goal of this study, it is a useful piece of software to build a direct photon identification algorithm.

First, it is helpful to understand the existing PHENIX algorithm. The algorithm takes the the center of a cluster and all electromagnetic calorimeter towers as input. It then takes the 49 towers around the center of the cluster and finds the center of energy. This is done by weighting each tower's position by the energy that has been deposited in that tower. The red dot in Figure 7.1 shows a possible position for the center of energy if the upper right of the 49 towers had more energy.

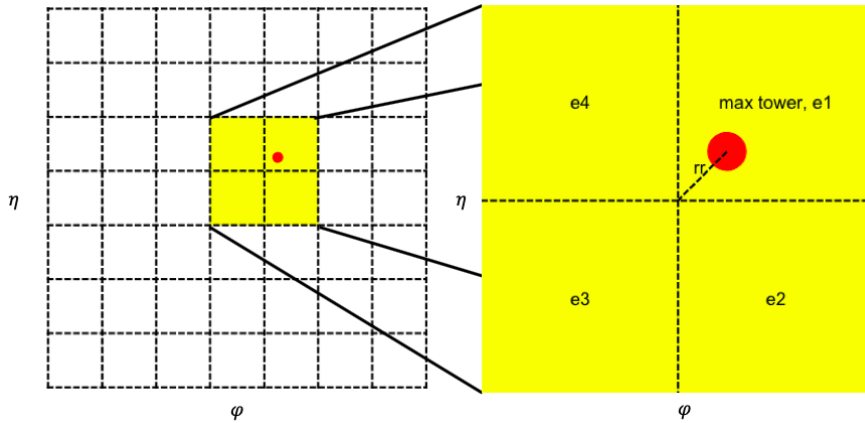


Figure 7.1: Diagram showing how the PHENIX photon/electron identifying algorithm extracts the variables to cut on.

The energies of the 4 towers that are closest to the center of energy are then assigned variables: e1 is the energy of the tower that the center of energy resides within, e2 is the energy of the tower at the same angle ϕ as e1, e4 is the energy of the tower with the same pseudorapidity as tower e1, e3 is the energy of the tower

farthest from the center of energy as shown in Figure 7.1. By comparing these values the algorithm calculates the distribution of energy in the central tower (e1), horizontally, vertically, in the offside tower (e3). Then using these extracted values, the algorithm compares to a template of what the values should be for a photon or electron. These values were generated from a combination of GEANT4 simulation and test beam studies.

7.2 Goals

An algorithm will be developed that is similar to the PHENIX algorithm. It will extract similar values to the PHENIX algorithm but instead of comparing to a template the values will be passed to a machine learning algorithm that will output the probability the cluster is from a direct photon. The goal of the algorithm is to be able to determine the probability a cluster is from a direct photon in both proton proton and gold gold collisions using calorimeter tower information.

7.3 Challenges

The background signal for this algorithm stems from the same sources as for the cluster isolation algorithm but offer their own unique challenges. Similar to the cluster isolation thermal photons and electrons background signals are fairly easy to remove; however, photons resulting from decays pose their own unique challenges. In this case only the calorimeter is being used to determine the identity of a cluster and a photon should produce a similar signal in the detector regardless of its origin. This poses a problem if a decay photon is to be distinguished from a direct one. The shape clusters in the electromagnetic calorimeter is first studied.

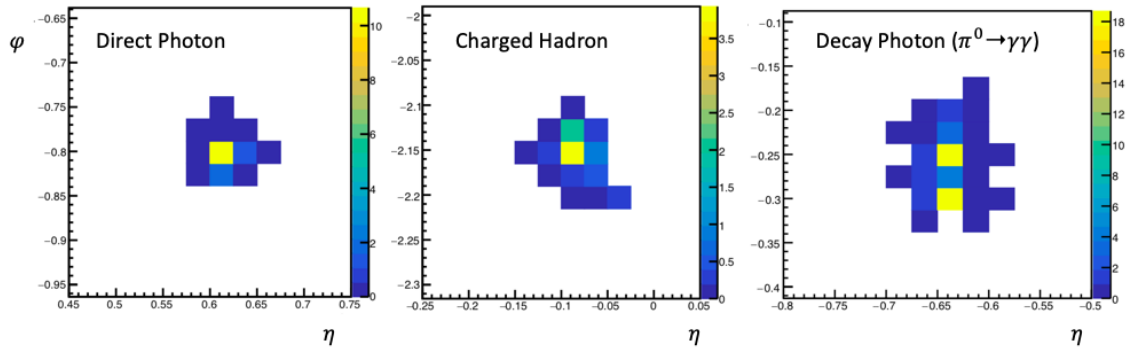


Figure 7.2: The energy distribution between towers of clusters from a direct photon, decay photon, and a charged hadron to show how different charged hadrons appear.

Figure 7.2 shows the energy distribution between towers in clusters resulting from a direct photon, a decay photon, and a charged hadron. The direct photon has almost all its energy deposited in the single tower, the hadrons energy is much more distributed, and the decay photon shows two points of large amount of energy being deposited. These examples indicate that a machine learning algorithm can separate these clusters based on their identity.

7.4 Training 1

With an indication that these clusters can be differentiated, an algorithm is constructed to extract similar variables to the PHENIX algorithm. Recall Figure 7.1 for the identity of $e1$, $e2$, $e3$, and $e4$.

$$e1t = \frac{e1 + e2 + e3 + e4}{E_{cluster}} \quad (7.1)$$

$$e2t = \frac{e1 + e2 - e3 - e4}{E_{cluster}} \quad (7.2)$$

$$e3t = \frac{e1 - e2 - e3 + e4}{E_{cluster}} \quad (7.3)$$

$$e4t = \frac{e3}{E_{cluster}} \quad (7.4)$$

First the algorithm was developed using proton proton collisions, much like the isolation cluster algorithm, due to the faster simulation times and easily managed background multiplicity. The algorithm was run on clusters of known origin to extract $e1t$, $e2t$, $e3t$, and $e4t$. Then every available TMVA machine learning algorithm was tested to see which can perform with the greatest efficiency while removing the most background. Four machine learning algorithms proved themselves to perform better than the others: MultiLayer Perceptron (MLP) neural network, Boosted Decision Trees (BDT), K Nearest Neighbor (KNN), and likelihood. After selecting primary candidates for the best machine learning algorithm they were trained with direct photons as the signal and a charged hadron as well as a decay photon background. Roc Curves are used to quantify the effectiveness of each of the algorithms with a given background. Figure 7.3 shows the performance of the 4 selected machine learning algorithms as well as the PHENIX algorithm. Recall that comparing to the PHENIX algorithm is not comparing to another direct photon identification algorithm, just a measure of the efficiency that is currently reached in identifying photons.

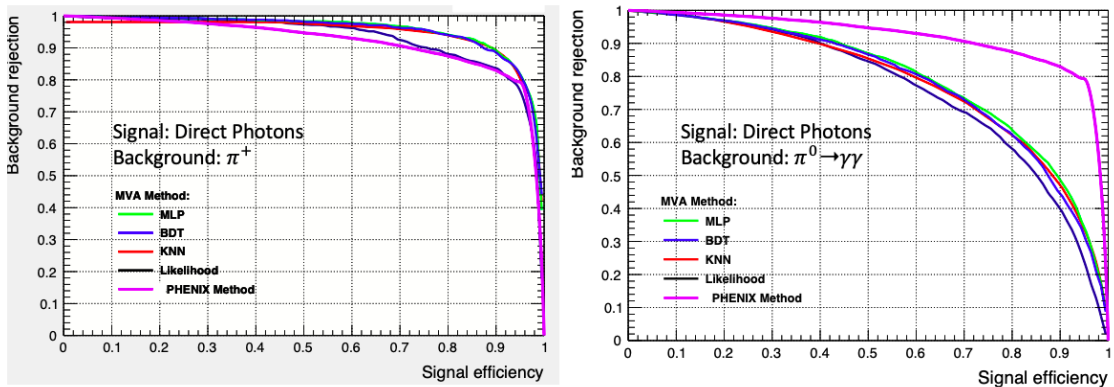


Figure 7.3: Roc Curves quantifying the effectiveness of several machine learning algorithms. Signal efficiency is what fraction of a sample of events is direct photons, background rejection is what fraction of the background signal is being removed from a sample of events. The $5 \text{ GeV} < E_T < 40 \text{ GeV}$ for signal and background clusters

The algorithm separates direct photon clusters from charged hadron clusters with a large efficiency and large background rejection, just as the PHENIX algorithm does; however, it does not attain as large an efficiency and background rejection with decay photons. Upon further investigation of the cluster shape of decay photons from π^0 s it was found that the shape of the resulting cluster depends heavily on the energy of the π^0 .

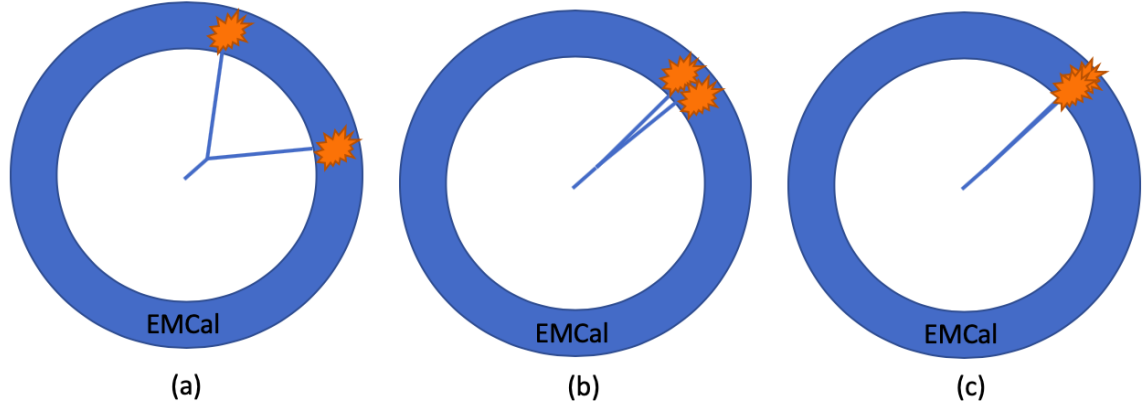


Figure 7.4: Diagram showing π^0 decays to 2 photons when the π^0 has approximately 5GeV (a), 15GeV (b), 30GeV (c).

In Figure 7.4(a) the photons would appear as a lone photon and may not be easily distinguished from direct photons; in Figure 7.4(b) the photon's clusters would contribute energy to each others cluster or they would be in the same cluster like the one shown in Figure 7.2 and be very distinguishable from a direct photon; in Figure 7.4(c) the photons merge into the same cluster, often even striking the same calorimeter tower, and be hard to distinguish from direct photons. In this case only the cluster isolation would separate direct photons from background.

TMVA does not train the machine learning algorithms in a way that this difference would be accounted for, as such a new method has to be developed to account for the difference in background signal depending on the photons energy. Either a different program for training machine learning algorithms could be used/developed, or alternatively several machine learning algorithms could be trained depending on the energy in the cluster being observed. Given the time constraints of this study the second option was pursued.

7.5 Training 2

Instead of training just 1 machine learning algorithm like in training session 1, 3 different versions of each machine learning algorithm were trained on different cluster energy ranges. As a start the three ranges of cluster energy were 5 – 15GeV, 15 – 25GeV, and 25 – 40GeV. The resulting algorithms were trained and tested, the resulting Roc curves showing the efficiency and background rejection of the algorithms is can be seen in Figures 7.5, 7.6, and 7.7. The machine learning algorithms trained on different energy ranges outperform the previously trained ones, allowing

between 40% and 80% background rejection with an 80% efficiency depending on the cluster energy.

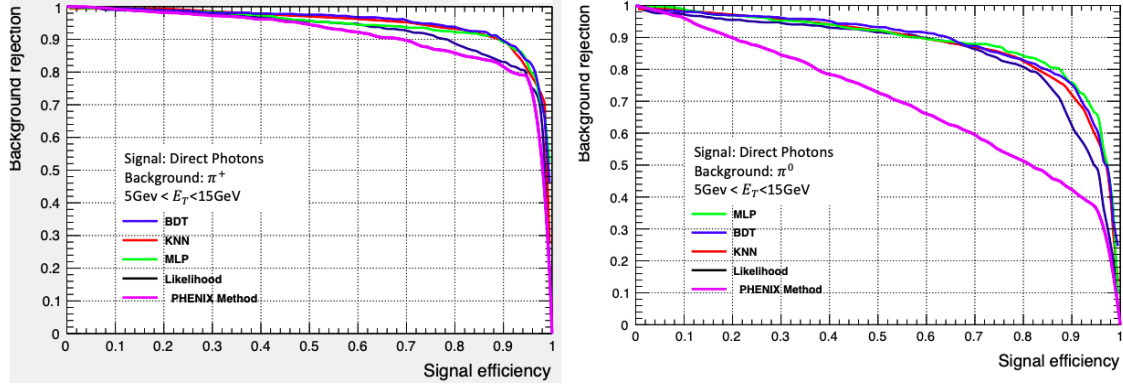


Figure 7.5: Roc Curves quantifying the effectiveness of several machine learning algorithms for clusters with 5 – 15GeV of energy

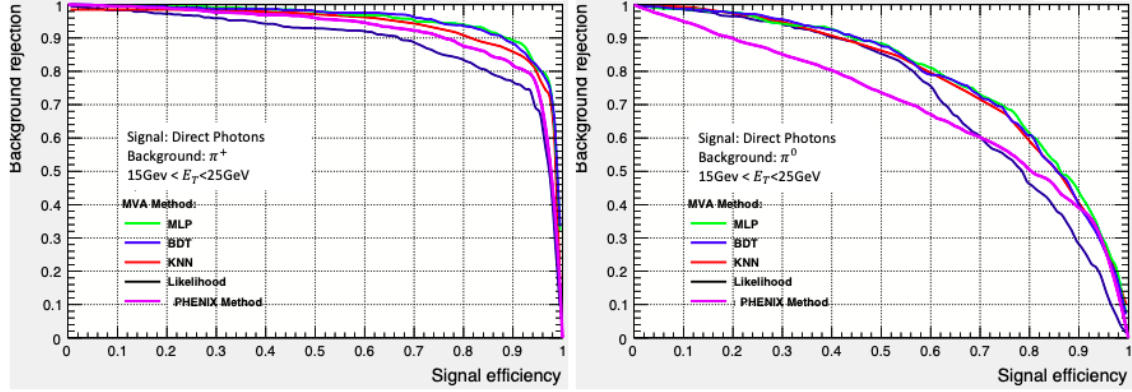


Figure 7.6: Roc Curves quantifying the effectiveness of several machine learning algorithms for clusters with 15 – 25GeV of energy

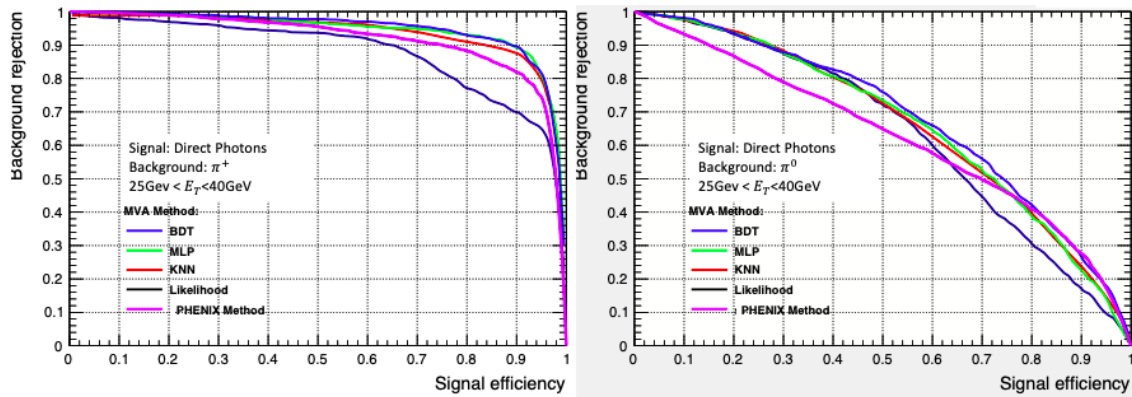


Figure 7.7: Roc Curves quantifying the effectiveness of several machine learning algorithms for clusters with 25 – 40GeV of energy

Of the 4 machine learning algorithms tested the MultiLayer Perceptron consistently outperformed the other 3 in efficiency and background rejection. Thus

the final photon identification algorithm takes a cluster's location and the detector calorimeter towers as input; extracts $e1t$, $e2t$, $e3t$, and $e4t$ from the 49 towers around the center of the cluster; passes these variables to the appropriately trained multiLayer perceptron depending on the clusters energy; and outputs the probability that the cluster was created by a direct photon.

7.6 Moving Forward

Now that direct photon identification algorithm has been proven to work with proton proton collisions, applying the same concept to gold gold collisions should prove trivial; however, at this point in the study the development of the algorithm was derailed. The sPHENIX software framework was updated and the resulting changes caused the code written for this study to no longer work as it was trying to reference code libraries that no longer existed.

To modify this direct photon identification algorithm to work with gold gold collisions one would first test the current algorithms performance with clusters from background subtracted gold gold events. After benchmarking the capabilities of the algorithm the next step would be to train the multilayer perceptron machine learning algorithm on the 3 different energy ranges and to test and tweak the algorithms to improve performance further.

Chapter 8

Conclusion

One of the main purposes of the proposed scientific program at sPHENIX is photon jet pair probes of Quark Gluon Plasma. In order to use photon jet pairs direct photons must be reconstructed and separated from various backgrounds. These backgrounds are made of decay photons and hadrons which begin showering in the electromagnetic calorimeter. In this study two approaches for separating signal from background were explored. The first method, cluster isolation, was shown to work in both proton proton and the subtracted background of gold gold collisions. The cluster isolation algorithm has been formatted into a package that functions as a part of the official sPHENIX Software Framework and can be used by any member of the collaboration. The second method, direct photon identification, is slightly more complicated. An option for identification is a method based off the existing PHENIX electromagnetic cluster identification algorithm. The ability of the existing PHENIX algorithm to separate signal from two backgrounds was benchmarked and a method which showed possible improvement using machine learning methods was developed. The direct photon cluster identification algorithm developed was shown to work in proton proton collisions with the opportunity to be applied to the high multiplicity environment of gold gold collisions.

Bibliography

- [1] A Adare, C Aidala, N.N. Ajitanand, Akiba, et al. Nonperturbative-transverse-momentum effects and evolution in dihadron and direct photon-hadron angular correlations in $p + p$ collisions at $s = 510$ gev. *Physical Review D*, 95, 04 2017.
- [2] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 2007.
- [3] Wit Busza, Krishna Rajagopal, and Wilke van der Schee. Heavy Ion Collisions: The Big Picture, and the Big Questions. *Ann. Rev. Nucl. Part. Sci.*, 68:339–376, 2018.
- [4] K. Adcox, S.S. Adler, M. Aizama, et al. Phenix detector overview. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(2):469 – 479, 2003. The Relativistic Heavy Ion Collider Project: RHIC and its Detectors.
- [5] Sarah Campbell. sPHENIX: The next generation heavy ion detector at RHIC. *J. Phys. Conf. Ser.*, 832(1):012012, 2017.
- [6] Yongsun Kim. The detector development and physics program in sphenix experiment at rhic. *Nuclear Physics A*, 982:955 – 958, 2019. The 27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions: Quark Matter 2018.
- [7] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [8] Xin-Nian Wang and Miklos Gyulassy. hijing: A monte carlo model for multiple jet production in pp , pA , and AA collisions. *Phys. Rev. D*, 44:3501–3516, Dec 1991.
- [9] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [10] Simon Turbide, Ralf Rapp, and Charles Gale. Hadronic production of thermal photons. *Phys. Rev.*, C69:014903, 2004.
- [11] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

All unaccredited diagram, figures, and plots were created using Microsoft Powerpoint, Mathematica, and Root by Chase Smith for this document.