STUDY OF JET QUENCHING IN QUARK GLUON PLASMA WITH MACHINE LEARNING

By

Lihan Liu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

August 12th, 2022

Nashville, Tennessee

Approved:

Julia Velkovska, Ph.D.

S. Victoria Greene, Ph.D.

Steven E Csorna, Ph.D.

Ran Tao, Ph.D.

ACKNOWLEDGMENTS

First and foremost, I'd like to express my sincere thanks to my parents, Zhian and Xinling, for all they have done for me. Because of their patience and love, I am simply a better man.

Second, I want to express my gratitude to my advisors, Marta Verweij and Julia Velkovska, for their guidance and support over the past several years. From them, I've learned so many skills for scientific research, which helps me to build a problem-solving mind. I also want to thank my Ph.D committee members, Prof. Victoria Greene, Prof. Steven E. Csorna and Prof. Ran Tao, for their mentorship and support. Thanks to the former and the current members of the Relativistic Heavy Ion Group at Vanderbilt University, as well as all graduate students of Physics and Astronomy department, for their help and encouragement. Special thanks to Dr. Shengquan Tuo for all your helps, I hope everything is going well with you and your families.

Third, I would like to thank my former mentor at University of Science and Technology of China (USTC), Prof. Zebo Tang, for his mentorship and encouragement. Thanks to all my friends for helping me survive anxieties and the pandemic.

Finally, thanks to my home country, China. I am feeling so lucky to have been born in a peaceful and prosperous era.

Thanks to the Advanced Computing Center for Research and Education (ACCRE) for providing computing resources. This work was supported in part by the US Department of Energy Grant No. DE-FG05-92ER40712, and by the Alfred P. Sloan Foundation, whose contributions are gratefully acknowledged.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Thesis Motivation

Quantum chromodynamics (QCD) is the theory of strong interaction, one of the four fundamental interactions with the others being gravity, electromagnetism and the weak interaction. QCD describes the interaction between quarks and gluons, which are the fundamental particles that make up composite hadrons such as the proton, neutron and pion. Most of the mass of the protons and the neutrons in everyday matter is in turn due to the binding energy of their constituent quarks as a result of the strong force between them.

The goal of high energy physics is to determine the most fundamental building elements of matter and to understand the interactions between these particles. The subject of high energy nuclear physics is the study of nuclear matter in extreme conditions, the phases in which it can exist, the transition between different phases, and characterizing the properties of the matter in each phase. To achieve these goals, hadron and nuclear colliders, such as the Large Hadron Collider (LHC) at CERN and the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven National Laboratory, are built as an experimental approach to unveil the mysteries in high energy particle and nuclear physics. Protons and nuclei are accelerated to speeds that are very close to the speed of light and collide with each other. These collisions produce massive particles that can be detected by detectors built around the collision points, providing a glimpse into what will happen in the aftermath of the collisions.

QCD predicts the existence of a deconfined state of matter, consisting of an extended volume of interacting quarks, anti-quarks and gluons, namely quark-gluon plasma (QGP). Calculations from lattice QCD have suggested a smooth crossover between hadrons at low temperature and QGP at high-temperature [1]. It is predicted that QGP comes into existence at temperatures and densities more extreme than any we know of in

the present natural universe, except perhaps in the interior of neutron stars, or in neutron star mergers [2, 3] Such extreme conditions are thought to have existed a few microseconds after the Big Bang. QGP can be created from heavy ion collisions at particle accelerators such as RHIC and LHC. While the details of QGP produced in heavy ion collisions, including its initial state effects, how the system attains thermal equilibrium and its dynamics are not yet very well understood.

The lifetime of QGP formed in heavy ion collisions is of the order of 10^{-22}s, making it challenging to probe the dynamics of the rich QGP matter. In ultra-relativistic heavy ion experiments, elliptic flow and jet quenching are arguably the two most important observables. Observation of an elliptic flow, almost as large as that predicted by ideal hydrodynamics, led to the claim of the formation of an almost perfect fluid at RHIC. The observation of jet quenching, i.e. the jet energy reduction in the presence of QGP, implies strong interaction between highly energetic partonic constituents and the dense colored medium.

In the early stages of a heavy ion collision, scatterings with large momentum transfer can happen between the partonic constituents of the incident nucleons resulting in highly energetic outgoing partons with sizeable transverse momentum (p_T). Further interactions of the scattered partons with the QGP may modify the spatial and momentum distribution of the final-state hadronic fragments relative to those in proton-proton collisions. This phenomenon, known as jet quenching, is useful in probing the properties of QGP. The properties of QGP can be accessed by studying the modification of the jet shower compared to the vacuum reference. As a common practice, pp collisions are regarded as a vacuum proxy, as no QGP is expected to be created, while in heavy ion collisions, jet variables are modified due to the interactions between the hard parton that initiates the jet shower and the partonic constituents of QGP.

Studies of jet substructure in heavy ion collisions aim to understand how the parton momenta and energies are distributed within the jet radius during the jet evolution with or without the presence of a QCD medium, which lies in the very forefront of investigation of the jet quenching phenomenon. Measurements of the jet

shapes and jet fragmentation functions have revealed modifications of the internal structure of jets due to quenching, providing opportunities for understanding in-medium parton evolution and medium response to hard probes.

Recent studies have reported successful measurements on the parton splitting functions in pp [4, 5] and in PbPb collisions [6]. In these studies, attempts are made to expose the "2-prong" internal structure of a single jet, such that the $1 \rightarrow 2$ parton splitting process can be studied. By removing wide-angle soft radiation until two subjets with sizable transverse momentum are found with the help of a jet grooming algorithm, the splitting of one parton into two well-separated partons is located. This hard splitting is subjected to medium-induced modifications, providing access to the properties of parton evolution with or without the presence of QGP. The measurements have suggested an enhancement of wider, and more unbalanced splittings due to quenching.

Modern data analysis techniques, especially machine learning, have attracted much attention these days. Machine learning is a data-driven method that automates analytical model building. Enthusiasts have been struggling with the application of machine learning in the study of jet physics and jet quenching. Various approaches have been developed with differences in the usage of artificial neural networks or in the training strategies.

This doctoral thesis focuses on the application of machine learning techniques in the study of jet quenching. The various aspects presented in the thesis are structured as follows. In Chapter 2, a brief introduction to quark-gluon plasma (QGP) is presented. Some experimental results are selected for the illustration of the jet quenching phenomenon. In Chapter 3, the full Monte Carlo event simulation procedures used in this study are described, with illustrations of the event mixing technique and background subtraction method. In Chapter 4, the state-of-the-art jet reconstruction and jet grooming methods are introduced, by which a general feature engineering method for jet studies is inspired. In Chapter 5, the full procedure of the machine learning

approach is introduced, involving the usage of an artificial neural network, its training, and its calibration. Finally, in Chapter 6, how the machine learning approach can be used in jet quenching studies are shown as results. In addition, the conclusion and overlook are also presented.

The PYTORCH framework [7] is used for the development of a machine learning application. The full code and data used in this study are available as open-source [8].

CHAPTER 2

The Quark Gluon Plasma

2.1 Quantum chromodynamics and quark gluon plasma

The strong interaction, dominating nuclear physics, is one of the four fundamental interactions, with the others being gravity, electromagnetism, and the weak interaction. The strong force takes effect at two different ranges. On a scale of about 1 to 3 fm, the effective field theory description involves the exchange of mesons (e.g., pions) that bind the protons and neutrons together to form the nucleus of an atom. On a scale that is smaller than the radius of a nucleon (about 0.8 fm), the relevant degrees of freedom are quarks and gluons, and the strong force confines the quarks within hadrons, such as the proton and neutron. Most of the mass of a proton or a neutron, the basic building blocks of everyday materials, results from the strong force field energy, while the individual quarks contribute only about 1% of the mass of a proton. On such a small scale, the strong force is mediated by the exchange of gluons, which are massless particles that act between quarks, antiquarks and other gluons. Quarks and gluons, which are often referred to as partons, carry color charges represented in three distinct colors, red, green and blue. Color charge is completely unrelated to the everyday meaning of color and is analogous to electromagnetic charge.

Details in the theory of quantum chromodynamics (QCD) describe the rules of quark-gluon interactions, the strength of which is characterized by the coupling constant α_s. In quantum field theory, the coupling constant depends on the four-momentum Q^2 transferred in the parton-parton scattering. For strong interactions, the Q^2 dependence is very strong between coupled gluons. A first-order perturbative QCD calculation gives

$$\alpha_s(Q^2) = \frac{12\pi}{(22-2n_f)\cdot ln(Q^2/\Lambda^2_{QCD})}, \tag{2.1}$$

where n_f is the number of quarks active to pair production (up to 6) and Λ_{QCD} is the fundamental energy scale ($\sim 0.22\,\text{GeV}$) in quantum chromodynamics. The dependence of a coupling constant on the energy scale is known as "running of the coupling".

The "running coupling constant" implies that for very small parton distances and high values of Q^2, the inter-quark coupling decreases, and vanishes asymptotically. In the limit of very large Q^2 (so-called the hard processes), quarks can be considered to be "free", known as asymptotic freedom. Therefore perturbative QCD calculation can be applied. On the other hand, at large distances, the coupling between the quarks increases, leading to the observed confinement of quarks and gluons in bound states, e.g., hardons. Consequently, at standard room temperature and density conditions, quarks and gluons are found in bound colorless states (hardons), either as color-anticolor ($q\bar{q}$) mesons or a color-triplet red-blue-green producing "white" (qqq) baryons. Measurements of the "running coupling constant" phenomenon are shown in Fig. 2.1.

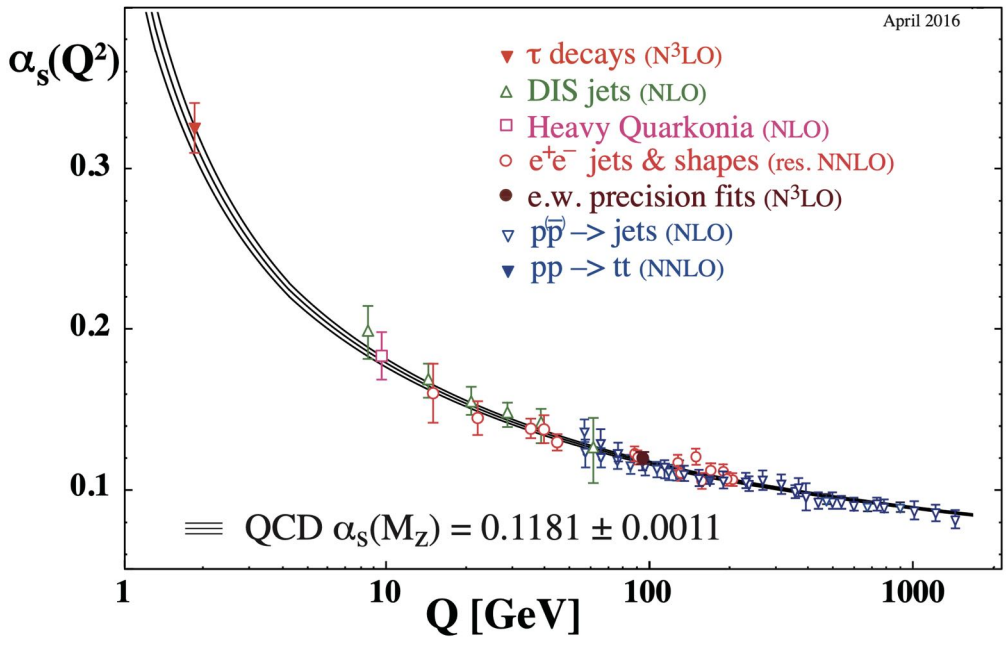

Figure 2.1: Summary of measurements of α_s as a function of the energy scale Q. Figure from [9].

Asymptotic freedom can be achieved by sufficiently increasing the energy density of a system, allowing for the deconfinement of quarks and gluons. Analogous to a plasma of ions, electrons and photons in quantum

electrodynamics, a phase of deconfined quarks and gluons at extremely high temperature and energy density is predicted to exist, known as the quark-gluon plasma (QGP). It is believed that QGP dominated the entire universe before hadronic matter was created. The necessary conditions of extremely high energy density for QGP formation are believed to have existed for a microsecond or so after the Big Bang. The study of QGP plays an important role in finite temperature QCD field theory, as well as understanding of the early evolution of the universe.

In the laboratory, QGP can be created copiously under controlled conditions in the collision of two heavy atomic nuclei accelerated to energies of more than a hundred GeV, which are able to achieve high enough energy densities to produce QGP and probe its transition through hadronic matter. The idea of exploring QGP dynamics using heavy ion collisions was first suggested by J. D. Bjorken in the early 1980s [10]. It is predicted that QGP could survive over a time scale $\gtrsim 5\,\mathrm{fm/c}$ at temperature of $\gtrsim 200-300\,\mathrm{MeV}$, achieved in relativistic nucleus-nucleus collisions. It was believed that depositing such a large amount of energy into a small volume would create the necessary conditions for the hadrons to melt into an assembly of weakly interacting quarks and gluons. Studies also laid out key experimental signatures of the short-lived QGP, including enhancements of strange particle production [11], higher rates of direct di-lepton and photon production compared to collision systems without the presence of QGP [12], melting of charmonia (bound states of charm-anticharm quarks, such as the J/ψ meson) due to color-screening in QGP [13], and the extinction of high-p_T jets due to parton energy loss in QGP [14].

2.2 Relativistic Heavy Ion Experiments

The first ultra-relativistic heavy ion experiments began in the mid-1980s with fixed-target experiments at the Alternating Gradient Synchrotron (AGS) at Brookhaven National Laboratory (BNL) and the Super Proton Synchrotron (SPS) at the European Center for Nuclear Research (CERN). At SPS, lead ions with beam energy up to 158 GeV/nucleon into gold or lead targets. Observations of a suppressed J/ψ meson (charmonium state)

yield by the NA50 experiment [15], strangeness enhancement measured by the WA57 experiment [16] among other results from SPS prompted the year 2000 announcement of a new state of matter, in which quarks and gluons are deconfined to "roam freely" [17]. With the start of operations of Relativistic Heavy Ion Collider (RHIC) in Brookhaven National Laboratory (BNL) the formation of a deconfined state of QCD matter was confirmed [18].

However, measurements from the RHIC experiments (BRAHMS, PHENIX, PHOBOS, and STAR) [19–22] contradicted earlier ideas that the medium produced in these collisions was a weekly interacting plasma of "free" quarks and gluons. Instead, clear signatures such as the collective behaviors of final-state particles, and the suppression of energetic particles, establishing that the QGP behaves more like a strongly-coupled near-perfect fluid of quarks and gluons flowing with the smallest specific viscosity ever observed for any fluid. Since then, experiments at RHIC have extensively characterized the medium properties and evolution of the QGP in a wide range of center-of-mass energy per nucleon pair ($\sqrt{s_{NN}}$) from 3 to 200 GeV. Since heavy ion collisions took place at the Large Hadron Collider (LHC) in 2010, results from four collaborations at CERN (CMS, ATLAS, ALICE and LHCb) have further advanced our understanding of the hot and dense QCD medium by focusing on QGP measurements at a much higher center-of-mass energy per nucleon pair ($\sqrt{s_{NN}}$) scales from 2.76 to 5.02 TeV [23].

2.3 Evolution of a heavy ion collision

To understand the thermodynamics properties of QCD matter under different conditions, a phase diagram is introduced, with temperature on the y-axis and baryochemical potential μ_b on the x-axis, as shown in Fig. 2.2. The baryochemical potential describes the excess of quarks over antiquarks in a system. Specifically, it refers to the energy needed to increase the system's baryon number by 1 unit. Everyday matter (nuclear matter) at standard room temperature and density resides within the hadron gas region in the phase diagram. The high temperature phase of QCD matter is in the form of QGP, while at low temperature and high μ_b a

color supercondicting phase is predicted [24, 25] that may possibly exists in the core of neutron stars, but this region of the phase diagram is not presently accessible experimentally.

The evolution of QGP formed in heavy ion collisions accessible at the LHC is believed to be similar to that of the early universe, starting from high temperature and $\mu_b = 0$. At RHIC, a large range of $\mu_b > 0$ can be investigated by lowering the beam energy from the top energy of $\sqrt{s_{NN}} =$ 200 GeV down to $\sqrt{s_{NN}} =$ 3 GeV. Perturbative QCD techniques, computational tools studying processes with large Q^2, break down when considering the bulk properties of the system across the phase diagram. However, lattice QCD calculations can help explore the behavior of a system of bound state hadrons with increasing temperature (at $\mu_b = 0$) and predict a phase transition from hadron gas to QGP. Lattice QCD calculations at finite μ_b are challenging [26], but the general expectation is that there is a first-order transition from hadron gas to QGP at high μ_b, while at low μ_b the transition is a smooth crossover [27]. Then, a big question emerges as whether a critical point, the endpoint on a phase equilibrium curve, exists or not. Locating the critical point on QCD phase space diagram is one of the primary goals of the beam energy scan (BES) program at RHIC, where collisions at center-of-mass energies ($\sqrt{s_{NN}}$) of 7.7 GeV, 11.5 GeV, 19.6 GeV, 27 GeV and 39 GeV are studied in BES phase 1 [28] while collisions at $\sqrt{s_{NN}}$ ranging from 3 GeV to 19.6 GeV are studied in BES phase 2 [29].

Lattice QCD calculations at $\mu_b = 0$ have suggested a sharp but continuous rise in energy density, pressure, and entropy divided by different powers of temperature when crossing the crossover band from the hadron gas phase to the QGP phase. This increase is indicative of the increasing number of the degrees of freedom in the system as the quarks and gluons become deconfined. The calculations also indicates a critical temperature, T_c, around 170 MeV, above which QGP can form.

A schematic view of the space-time evolution of a heavy ion collision is presented (Fig. 2.3) with colors indicating the temperature of the plasma formed, which also suggests that QGP forms later with further distance (bigger z or rapidity y) away from the collision point. Two nuclei coming from opposite beam

[2]Figure from NSAC 2015 Long Range Plan for Nuclear Physics.

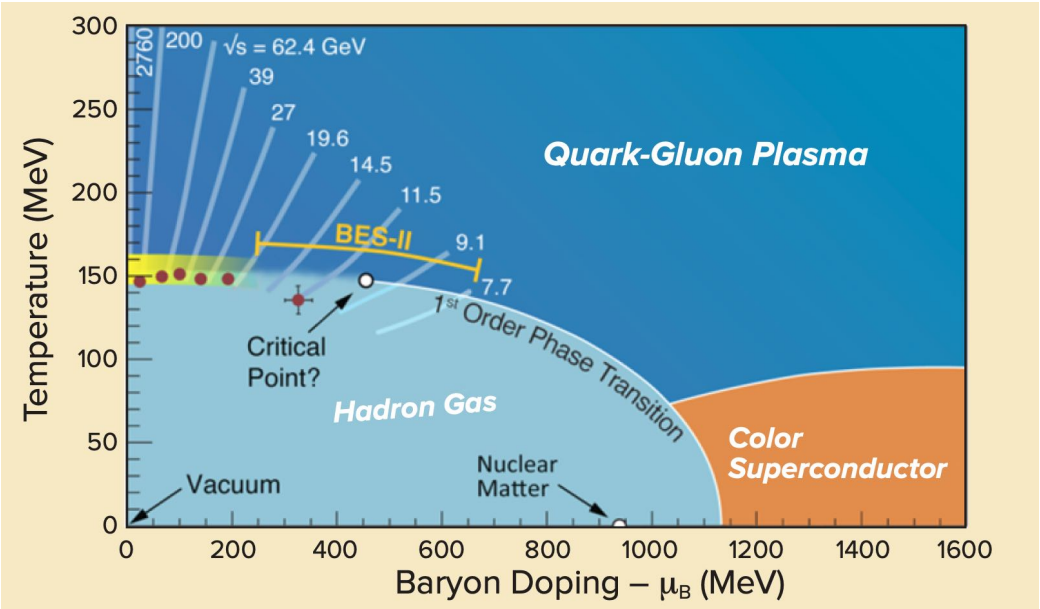

Figure 2.2: QCD phase space diagram as a function of baryochemical potential μ_b and temperature T. The transition between QGP and hadron gas is represented as a continuous crossover at low μ_b with a yellow band [2].

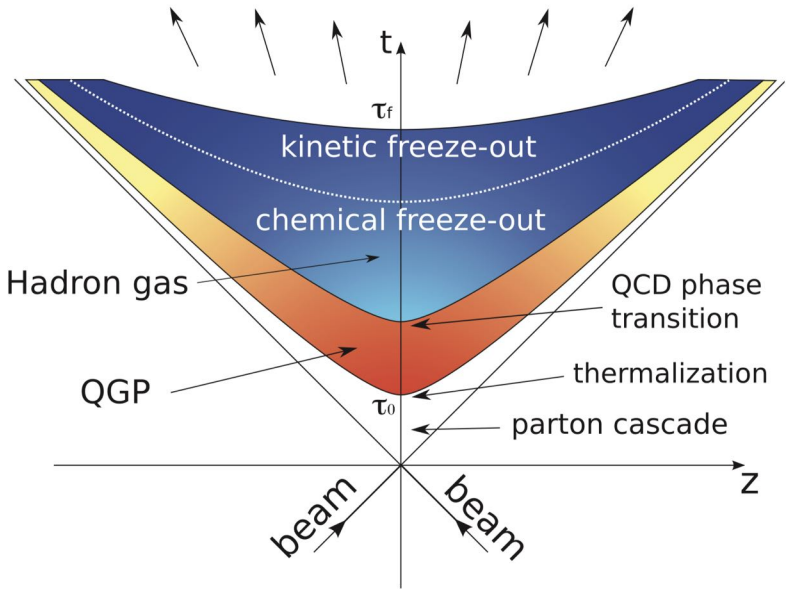

Figure 2.3: Space-time evolution of a heavy ion collision. Figure from [30].

directions come into contact at $t = 0$ and $z = 0$, liberating many of the partons as a result of the inelastic collisions between individual nucleons. The frequent scattering of the partons leads to the local equilibrium at time τ_0, after which the relativistic hydrodynamics can be used to describe the evolution of the system. Then, the QGP expands and cools until its temperature drops to some point around the critical temperature (T_c), after which hadrons come into presence. The hadrons may interact inelastically and change their identities before the chemical freeze-out temperature (T_{ch}) is reached. Hadrons may still interact with each other in elastic collisions and exchange momenta and energies. Finally, as the temperature drops to the kinetic freeze-out temperature (T_{fo}) at time τ_f, hadrons stop interacting and freely stream outwards. Signals from final-state particles are collected by detectors that are built around the collision point, which are then used for further analysis.

2.4 Event kinematics and centrality

A few basic phenomenological observables that are commonly used in heavy ion collisions are briefly discussed in this section. As a convention, natural units are adopted in high-energy physics where $\hbar = c = 1$. This implies that quantities such as the mass, energy, momentum and temperature are presented in the unit of electronvolt (eV). With natural units, the relativistic energy-momentum equation, which describes how total energy scales with the rest mass and momentum, is

$$\begin{aligned} E^2 &= p_x^2 + p_y^2 + p_z^2 + M^2 \\ &= p_T^2 + p_z^2 + M^2, \end{aligned} \tag{2.2}$$

where p_T (transverse momentum) is the portion of momentum transverse to the z-axis. Conventionally, the beam direction is defined as the z-axis and is perpendicular to the transverse plane spanned by the x-axis and the y-axis, as shown in Fig. 2.4 (left). Detectors are usually built in a cylindrical shape, covering full or part of the azimuthal angle (ϕ) around the beam direction.

In special relativity, rapidity (y) is commonly used as a measure for relativistic velocity,

$$y = \frac{1}{2} ln(\frac{E + p_z}{E - p_z}). \tag{2.3}$$

In an experiment often the mass of the particle is not determined and the rapidiy can not be measured. In this case, the pseudorapidity (η) is used. It is defined as

$$\begin{aligned} \eta &= \frac{1}{2} ln(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z}) \\ &= -ln[tan(\frac{\theta}{2})], \end{aligned} \tag{2.4}$$

where θ is the polar angle of an outgoing particle. The relation between the polar angle θ and pseudorapidity η is shown in Fig. 2.4 (right) for clarity. For particles moving with relativistic velocities close to the speed of light (c) rapidity and pseudorapidity differ very slightly. A second-order Maclaurin expansion gives that

$$y \approx \eta - \frac{cos\theta}{2} (\frac{m}{p_T})^2, \tag{2.5}$$

which clearly shows that at large $p_T \gg m$, pseudorapidity becomes equal to (true) rapidity.

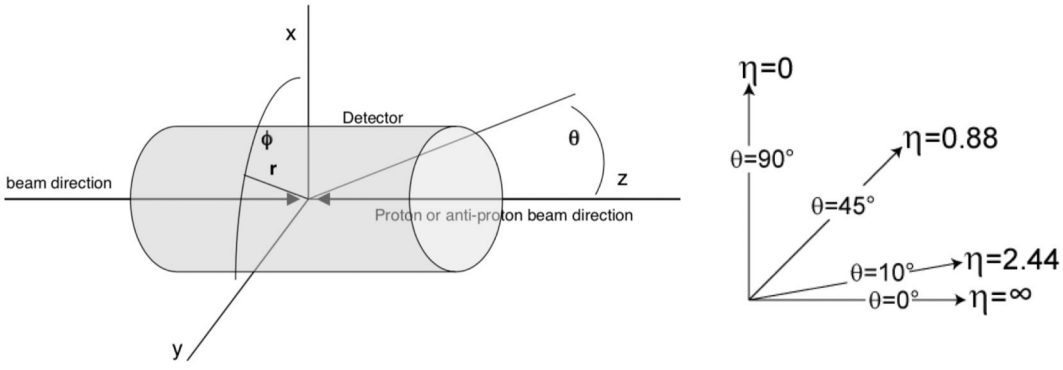

Figure 2.4: Left: Cylindrical coordinate system used in high-energy collider physics measurements. Right: Relation between polar angle θ and pseudorapidity η. Figures from [31].

In heavy ion collisions, the impact parameter (b) of the incoming nuclei significantly affects the physics of the ensuing processes due to the fact that it controls the fraction of overlap between two Lorentz-contracted

"discs", thus determining the initial geometry and the size of QGP. Collision events may vary from being entirely head-to-head (most central) to barely missed (most peripheral) as they approach each other. Any event in between corresponds to an elliptic overlap region controlled by the impact parameter b. The impact parameter is difficult to measure, with exceptions in Monte Carlo simulation as it is programmable. In practice, the degree of overlap or the "centrally" of a collision is described as a percentage quantile of the total nuclear inelastic cross section, ranging from 0% for the most central case to 100% being the most peripheral case. When centrality comes to the selection of samples, the $0-10\%$ centrality class refers to the top 10% events with the smallest impact parameters (biggest overlap sizes).

Reliable control on the impact parameter can be achieved on an event-by-event basis, as it can be inferred from experimental observables. It is assumed that in a collision event, the impact parameter, the number of participating nucleons (N_{part}) and the number of binary collisions between the incoming nucleons (N_{coll}) are monotonically related to observables, such as the total number of produced particles (multiplicity) and/or the energy deposited in detectors. Theoretical calculations from a "Glauber model" [32] can describe the relationship between these quantities and the collision centrality, shown in Fig. 2.5.

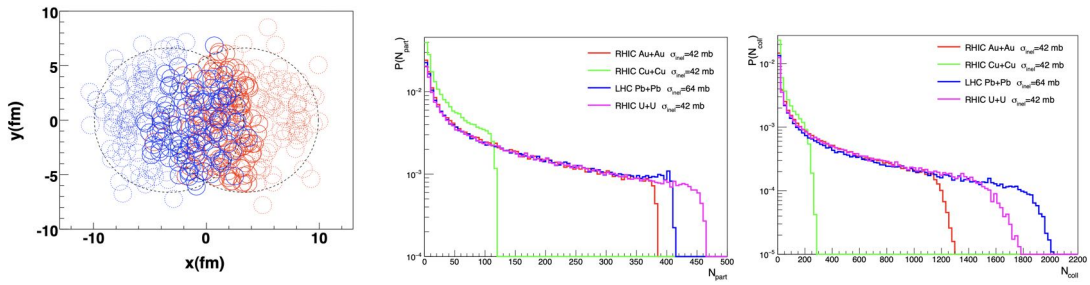

Figure 2.5: Left: Simulation of a PbPb collision at LHC with an impact parameter $b \approx 7$ fm. Participant nucleons (N_{part}) are shown with solid circles and spectators with dotted circles. Middle and Right: Distributions of N_{part} and N_{coll} in different collision systems at RHIC and LHC according to Glauber model calculations. Figures from [33].

2.5 Jet and Jet quenching

Most of the interactions in relativistic heavy ion collisions are soft processes, in which the momenta transferred between the interacting partons are relatively small. While very rare, hard processes do happen where a large amount of momentum can be transferred, producing a pair of back-to-back high p_T partons (quarks and gluons), or pairs of partons and electroweak bosons (γ and W/Z bosons) that appear at a large angle relative to the beam direction (see Fig. 2.6). Considering that hard scattering processes occur prior to QGP formation, the pair of the outgoing products undergo interaction with the surrounding medium if there is any, followed by fragmentation and hadronization. As a result, detected final-state particles associated to the hard scattering process, known as hard probes, play an important role in probing QGP. A highly energetic parton coming from the hard scattering process produces a cone-shaped spray of final-state particles, the collection of which is referred to as a "jet". Fully reconstructed jets are observed to connect to the energy and direction of the initial hard-scattered high p_T partons, providing crucial information about the organization and properties of QGP.

As a high-p_T parton passes through the strongly coupled QGP, it: (i) loses energy and longitudinal momentum, (ii) gains momentum transverse to its original direction, (iii) deposits energy and momentum into the droplet of QGP. These phenomena, unique to the case where a QCD medium is present, lead to modifications on the total energy of a jet and on how the energy is distributed within the jet's cone, compared to that in the cases where QGP is absent. Such modifications, widely known as the "jet quenching" effects, offer opportunities to further understand the quark-medium interaction mechanism in the QGP.

Energetic partons may lose energy through both elastic and inelastic collisions with the constituents of the QGP medium, with their typical diagrams shown in Fig. 2.7.

The energy loss though inelastic collisions, so-called radiative energy loss, is often considered as the most important component in studying the parton energy loss in a QCD medium. Even in vacuum, energetic

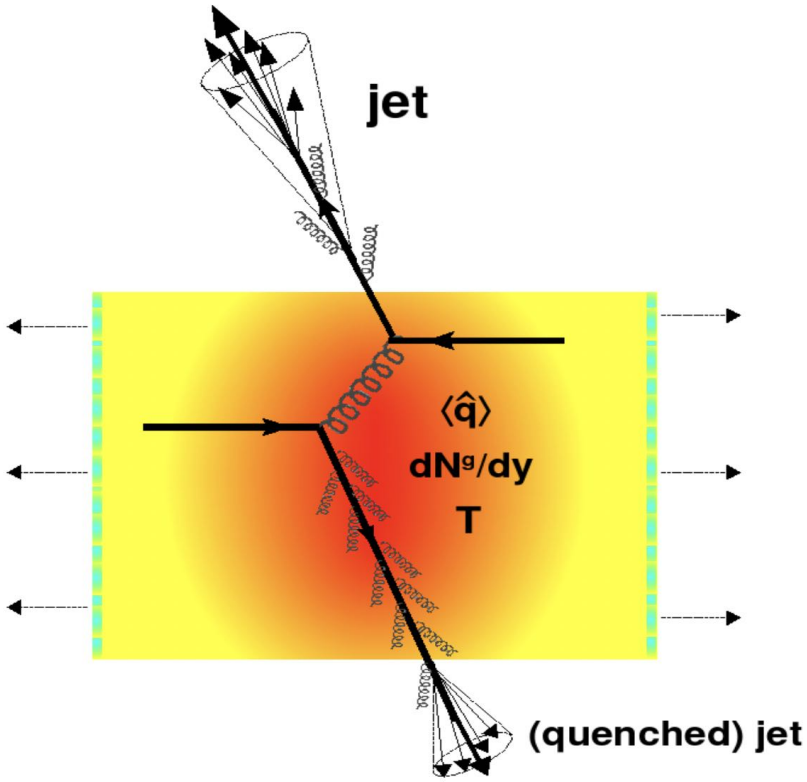

Figure 2.6: "Jet quenching" in a head-on nucleus-nucleus collision. Two quarks undergo hard scattering: one goes directly into vacuum, the other travels through the medium and loses energy. Medium properties such as transport coefficient $\hat{q}$, gluon density dN^g/dy and temperature T affect the energy loss process. Figure from [34].

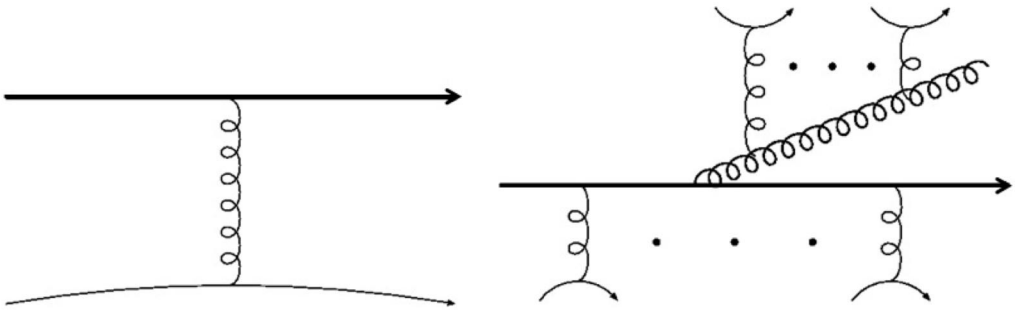

Figure 2.7: Typical diagrams for parton energy loss in a QCD medium though elastic (left) and inelastic (right) collisions. Figure from [35].

partons produced in hard scattering will under go vacuum splitting process, while the presence of a QCD medium will lead to modifications on the gluon radiation or the parton splitting process. Theoretical approach to the energy loss in a QCD medium often treats medium partons as scattering centers, and the emitted gluons are considered to be able to maintain coherence to its emitter. This leads to an estimation of the total energy loss for a parton traveling in a QCD medium [36] as,

$$\Delta E \sim L^2, \tag{2.6}$$

where L is the path length through the medium. Such path-length dependence is different from that of the bremsstrahlung in QED, which is not squared, suggesting that parton energy loss in QCD medium might be large.

The energy loss occuring in the $2 \rightarrow 2$ elastic collisions is usually called elastic or collisional energy loss, which is considered to be small for light flavor partons, especially when the energy of the jet is sufficiently high. On the contrary, for heavy flavor partons or at low and intermediate energy regimes, the collisional energy loss contributes dominantly to the total energy loss. The collisional energy loss, though relatively small for light flavor energetic jets, may still play an important role in understanding the nuclear modification on the full jet in QGP. It is also an essential ingredient in studying the medium response to the propagating jet, considering the fact that scattered medium partons may gain energy and momentum from parton collisions ("recoil" effects).

2.6 Experimental Observations

From an experimental perspective, QGP is defined as a locally thermalized state of matter in which quarks and gluons are deconfined. The color degrees of freedom become manifest over the nuclear, rather than merely nucleonic, volumes. The thermalization and the deconfinement mark the formation of QGP in heavy ion collisions. In this section, an overview of experimental observations on QGP is presented, with

selected contents including measurements on both collective and localized behaviors of QGP, specifically the flow and the jet quenching phenomena.

2.6.1 Bulk Properties

Bulk properties are properties due to interactions between system participants, which are referred to as quarks and gluons as they become deconfined in QGP. Explorations of its bulk properties have led to discoveries about the freeze-out temperatures, specific viscosity and anisotropic flow, which provide information about the degree of thermalization and the equation of state (EoS) related to QGP formation. In experiments, these discoveries are made with measurements of particle multiplicities, total yields of different particle species and their momentum spectra, especially at low transverse momentum (p_T), where most of the particles are produced in heavy ion collisions.

2.6.1.1 Strange particle production and chemical freeze-out

In heavy ion collisions, inelastic collisions between particles stop at a certain temperature, the chemical freeze-out temperature, and the abundance of system elements becomes fixed. Measurements on the yields of different hadron species, especially strange particles, provide information about the bulk properties of QGP at the chemical freeze-out temperature.

Due to the fact that the mass of a strange quark ($\sim 150\,\mathrm{MeV}$) is similar to the critical temperature of QCD transition, the yield of strange particles is expected to be enhanced in heavy ion collisions, which serves as a signature of QGP formation. The p_T integrated yields of primary strange particles (K^0_s, Λ, $\bar{\Lambda}$) and multi-strange particles (Ξ^-, $\bar{\Xi}^+$, Ω^-, $\bar{\Omega}^+$) are compared with to those of non-strange particles such as pions [37], as shown in Fig. 2.8. The measured ratios between strange particle yields and that of pions have suggested that strange quarks are equilibrated in the system, with exceptions of the most peripheral heavy ion collisions.

Interestingly, these ratios have shown a significant dependence on the total multiplicity instead of the initial volume (collision centrality) or the initial energy density (collision energy).

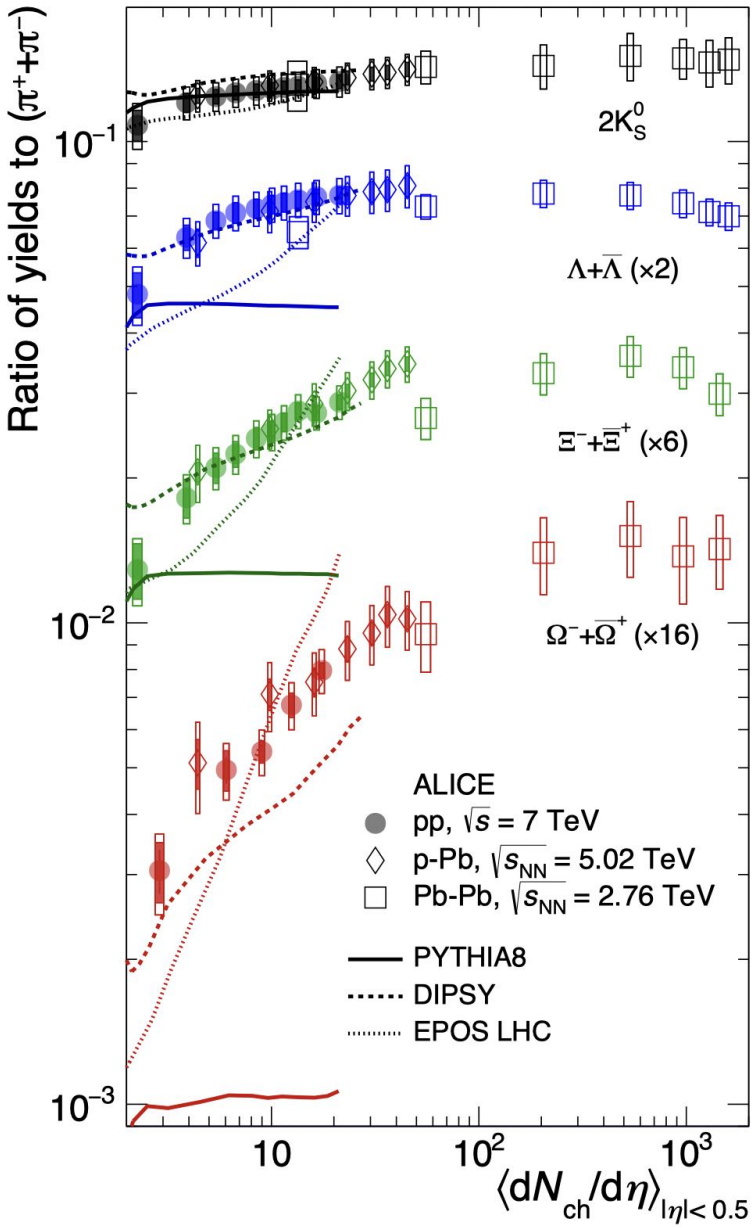

Figure 2.8: p_T-integrated yield ratios to pions ($\pi^+ + \pi^-$) as a function of $\langle dN_{ch}/d\eta \rangle$ measured in $|y| < 0.5$.

Such particle production ratios reflect how far a system is away from chemical equilibrium. The observation of strange particle saturation in high multiplicity PbPb events leads to a common belief that QGP is formed and reaches chemical equilibrium. Though pp events are conventionally regarded as systems where no QGP is formed, those high multiplicity pp events have shown enhanced strange particle yields, which is

also the case in pPb collisions. The similarities of strange particle production in pp, pPb and PbPb collisions are understood to be connected to the formation of a deconfined QCD phase at high temperature and energy density.

2.6.1.2 Spectra and kinetic freeze-out

The inelastic collisions between hadrons cease after the temperature drops down to the chemical freeze-out temperature. Even though the particle species are fixed, hadrons still interact with each other and exchange momentum before the temperature reaches the kinetic freeze-out temperature (T_{fo}). Measurements of hadron transverse momentum spectra provide information about the characteristics of the system at kinetic freeze-out. To better understand the transverse expansion of the system, a hydrodynamic-motivated fit method [38] is introduced to the hadron p_T spectra, which extracts a random motion component and a collective motion component, which are respectively described by fit parameters of the kinetic freeze-out temperature T_{fo} and the radial flow collective velocity $\langle\beta_T\rangle$.

Figure 2.9 shows the χ^2 contours of extracted fit parameters T_{fo} and $\langle\beta_T\rangle$ for different hadrons at different centrality bins, measured by the STAR experiment [22]. For copiously produced hadrons (π, K, p) in AuAu collisions, as the collisions appear more and more central, the extracted T_{fo} becomes smaller, while the $\langle\beta_T\rangle$ becomes larger. This implies that compared to the most peripheral collisions, systems in the most central collisions tend to be more difficult to reach the kinetic freeze-out, thus having a lower T_{fo}. The systems in central collisions develop stronger radial flow, leading to larger $\langle\beta_T\rangle$ than that in peripheral collisions. Interestingly, the multi-strange particles such as ϕ and Ω appear to have a higher freeze-out temperature and a lower radial flow velocity than that of π, K, p, which indicates that those multi-strange particles undergo little interactions at hadronic level with the expanding matter. If that is true, multi-strange particles should accumulate radial flow velocity before chemical freeze-out, leading to their sensitivity to the partonic stage of the QCD matter.

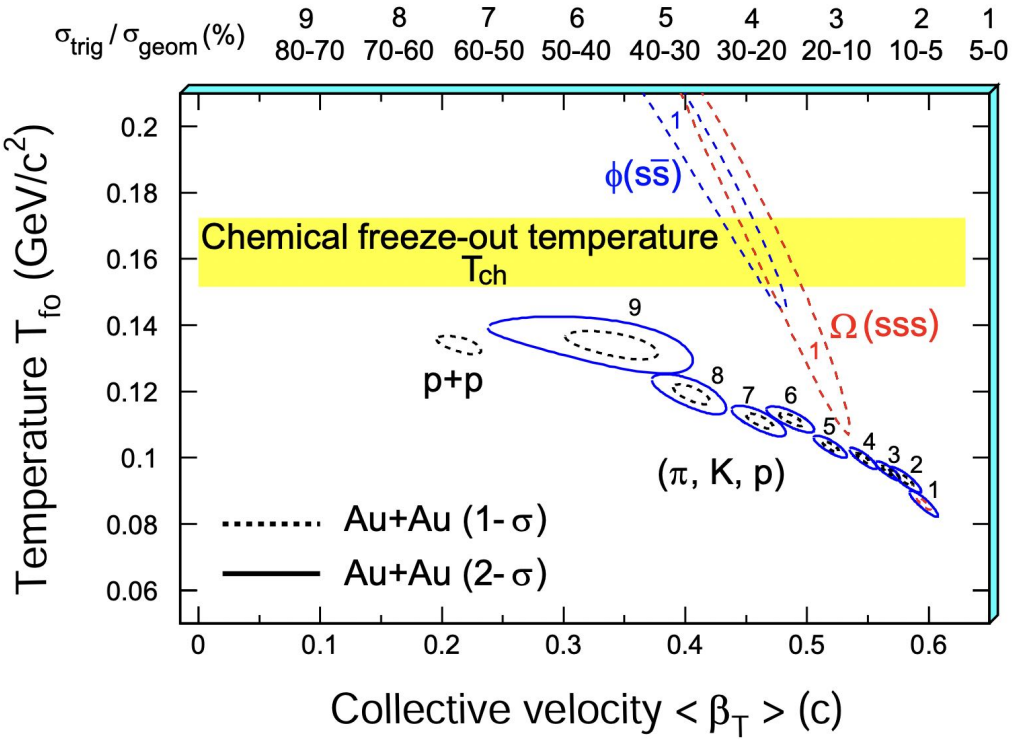

Figure 2.9: χ^2 contours for T_{fo} and $\langle\beta_T\rangle$, extracted from thermal + radial flow fits, for copiously produced hadrons (π, K, p) and multi-strange hadrons (ϕ, Ω) with 9 different centrality selections.

2.6.1.3 Collectivity in QGP

Due to the strongly coupled nature of an almost hydrodynamically ideal QGP liquid, any spatial anisotropy in the initial stage of heavy ion collisions could manifest itself in the final state. The anisotropy could arise from either the elliptical overlap region of colliding nuclei or the local energy hot spot in the incoming nuclei. The non-circular geometry of QGP formed in heavy ion collisions, especially in off-central collisions, as shown in Fig. 2.10 (left), would produce initial azimuthal anisotropies. If the QGP were a weakly coupled gas of quarks and gluons, the initial anisotropies would be washed out and would result in an isotropic distribution of final-state particles. However, for a strongly coupled liquid such anisotropies translate into a fast hydrodynamic expansion with anisotropy in pressure gradients. The anisotropies are minimally affected during the thermalization and the hydrodynamization stages, and survive through the hadronization and the free-streaming stages [39]. Thus, the collective behaviors of final-state particles, also known as the "flow", can be observed.

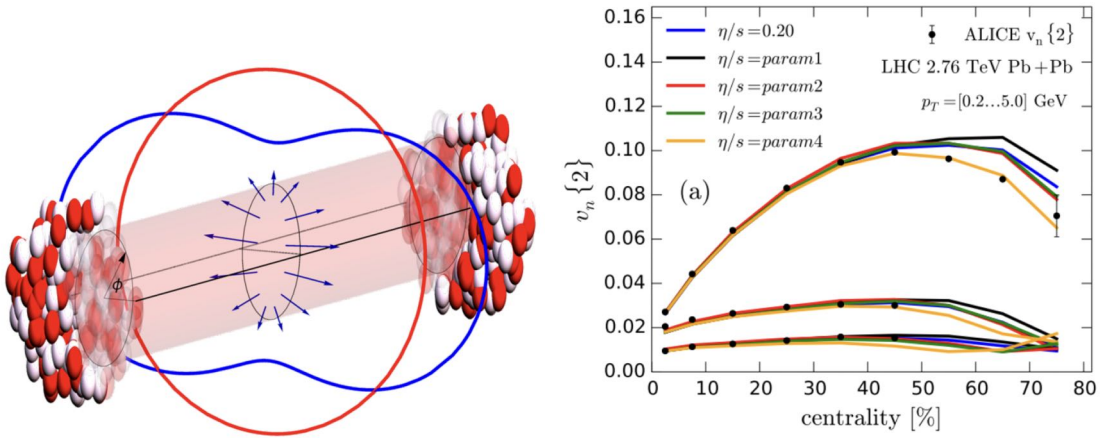

Figure 2.10: Left: Example of an off-central heavy ion collision which produces an approximately elliptical overlap region, resulting in anisotropy gradients in the azimuthal plane. Right: A hydrodynamic model with temperature-dependent parameterization of η/s is compared with ALICE measurements of the Fourier coefficients v_n (n=2 to 4 from top to bottom). Figures from [40].

In order to quantify the measurement of azimuthal anisotropy, a Fourier expansion can be performed on

the angular distribution of (charged) hadrons registered in the final state of the collisions

$$\frac{dN}{d\phi} \propto 1 + \sum_n 2v_n(p_T)cos(n(\phi - \varphi_{RP})), \tag{2.7}$$

where ϕ is the azimuthal angle of a particle relative to the beam direction, N is the total number of particles in the event, φ_{RP} is the azimuthal angle of the reaction plane and v_n is the n-th order harmonic coefficient. The different harmonic coefficients represent different aspects of the global flow behavior. The first order coefficient v_1 quantifies the flow of particles in the plane that is traverse to the beam direction, known as "directed flow". The second order v_2, so-called the "elliptic flow", is the largest component that characterizes the anisotropy due to the elliptic overlap region of the colliding nuclei. Higher order coefficients, for example, the third order v_3 known as the "triangular flow", are results of fluctuations or lumpiness in the initial energy density in the area of overlap between the colliding nuclei.

Figure 2.10 (right) shows ALICE measurement on the flow harmonic coefficients with hydrodynamic fits [41, 42]. The elliptic flow, parameterized by v_2, is observed to be strongest in mid-central collisions $(40-50\%)$ due to the eccentric shape of the overlap region between incoming nuclei. While v_2 is observed to be small in the most-central collisions $(0-5\%)$ due to the circular overlap region. The v_n measurements, in comparison to calculations from a hydrodynamic model with temperature-dependent parameterization of η/s, provide constraints on the viscosity of the QCD medium.

Studies have been done on both the elliptic and the triangular flow, in addition to their relations with the initial-state geometry. A geometry scan experiment in small collision systems, $p+Au$, $d+Au$ and ^3He+Au, are conducted at RHIC. The Figure 2.11 shows the hydrodynamic evolution of a typical $p+Au$ (up), $d+Au$ (mid) and ^3He+Au (bottom) collision, calculated from a hydrodynamic model. In the first panel of each row, regions marked in red indicate high energy density due to geometry overlap. The following three panels show snapshots of the temperature distribution at three successive time points. The initial energy configuration sets a pressure gradient field, which translates into a velocity field that determines the azimuthal momentum

distribution of final-state particles.

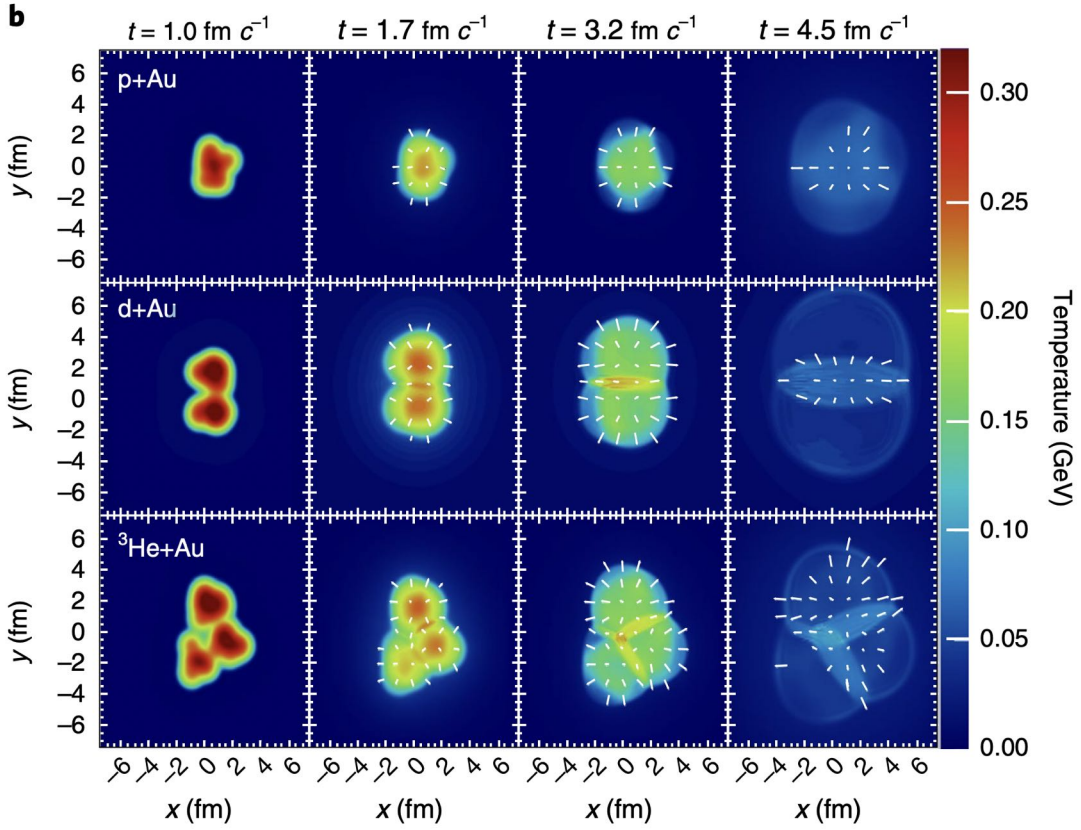

Figure 2.11: Evolution of a typical $p+Au$ (up), $d+Au$ (mid) and ${}^3He+Au$ (bottom) collision, calculated from a hydrodynamic model. Figure from [43].

The PHENIX experiment measured both v_2 and v_3 in the three small collision systems [43], shown in Fig. 2.12, providing evidence for the origin of the collective flow at the smallest scales. The measurement has shown the orderings of v_2 and v_3,

$$\begin{aligned} v_2^{p+Au} &< v_2^{d+Au} \approx v_2^{^3He+Au} \\ v_3^{p+Au} &\approx v_3^{d+Au} < v_3^{^3He+Au} \end{aligned} \tag{2.8}$$

rooted in the initial geometric eccentricities. This ordering implies that the hydrodynamic evolution can efficiently translate the initial geometry eccentricities into the collective behavior of final-state particles. It also suggests that QGP forms in such small collision systems.

In addition to measurements of the flow harmonic coefficients, two- and multi- particle correlations are

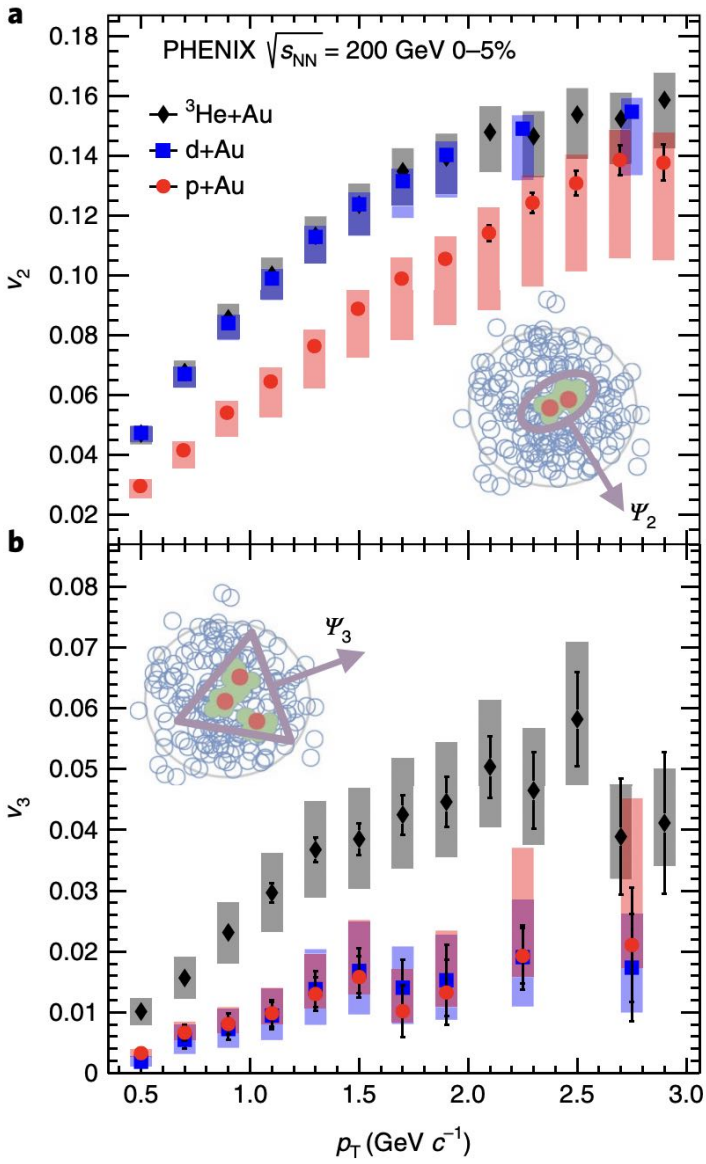

Figure 2.12: PHENIX measurements of v_2 (a) and v_3 (b) in $p+Au$, $d+Au$ and ^3He+Au collision systems. Figure from [43].

also utilized to quantify the collective behaviors of particles. In the studies of two-particle correlations, the relative pseudorapidity ($\Delta\eta$) and the relative azimuthal distance ($\Delta\phi$) are measured between a pair of particles, a trigger particle and an associated particle in the same event. Pairs of particles are selected within the same p_T range. Two-dimensional histograms of $\Delta\eta$ versus $\Delta\phi$ are shown in Fig 2.13 for pp and PbPb collisions, measured by the CMS experiment [44, 45]. An away-side ridge, referring to $\Delta\phi \approx \pi$, is expected due to the conservation of momentum. The hadrons comprising dijet products of hard scattering emerge back-to-back in the azimuthal angle, but they are not necessarily at the same rapidity. In the two-particle correlation pattern the particles from the back-side jet swing at a different η in each event leading to a smeared peak as a function of η at the away-side, when many events are are overlayed. In pp collisions, a jet-like peak is observed at the near-side $\Delta\phi \approx 0$, due to the showering process from a high-p_T parton coming from hard scattering, while in PbPb collisions, a near-side ridge is also observed, showing collective long-range correlations between the particles due to flow. The 2D histogram is projected onto the $\Delta\phi$ axis over the range outside of the jet peak $|\eta| > 2$, to which a Fourier function fit can be performed in order to extract the flow harmonic coefficients v_n.

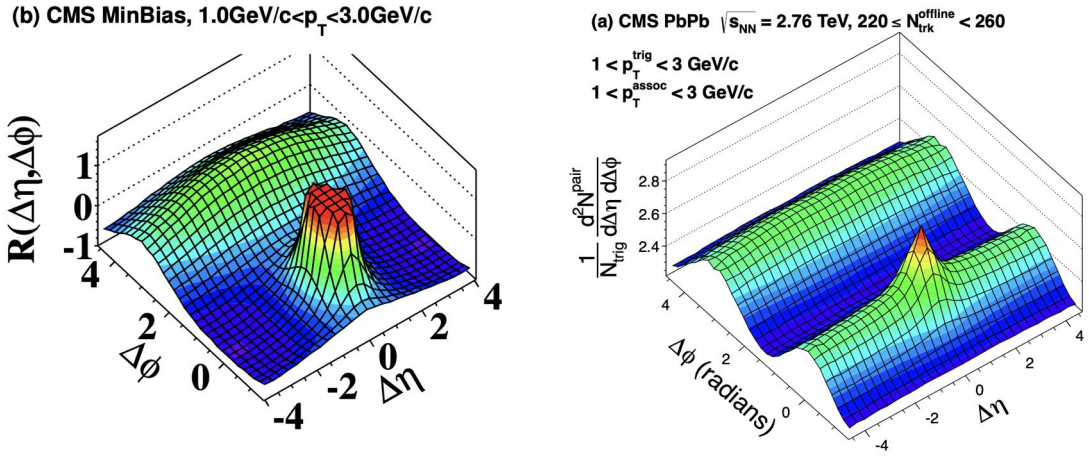

Figure 2.13: 2D maps of dihadron correlations for pp (left) and PbPb (right) collisions, measured by the CMS experiment. Figures from [44, 45].

2.6.2 Hard Probe: Jet

2.6.2.1 R_{AA} and dijet asymmetry

A wide variety of measurements on jet observables have been performed in PbPb collisions and in pp collisions. Jets produced in pp collisions are considered as vacuum references due to the absence of the QGP. The inclusive jet spectra have been measured by the ATLAS experiment, which has shown a significant suppression in inclusive jet production in PbPb collisions compared to that in pp collisions. Nuclear modification factor (R_{AA}), the observable that is used to quantify the suppression, is defined as

$$R_{AA} = \frac{\frac{1}{N_{evt}} \frac{d^2 N_{jet}}{dp_T dy}|_{cent}}{\langle T_{AA} \rangle \frac{d^2 \sigma_{jet}}{dp_T dy}|_{pp}}, \tag{2.9}$$

where N_{jet} is jet yield in PbPb collisions, σ_{jet} is jet cross-section in pp collisions, N_{evt} is the total number of collisions within centrality range, $\langle T_{AA} \rangle$ is the mean nuclear thickness accounting for geometric enhancement. Interestingly, R_{AA} remains below 1 even for high-p_T jets, indicating a considerable energy loss due to jet medium interactions.

The total energy loss experienced by a propagating parton depends on the path length in QGP. A pair of back-to-back partons originating from hard scattering may travel through QGP with different path lengths, leading to momentum asymmetry in associated full-reconstructed jets. For illustration, the dijet asymmetry in a single PbPb collision is shown in Fig. 2.15, which clearly shows an asymmetry between the transverse momenta of the leading and the subleading jet.

The dijet asymmetry ratio (A_J), is defined in Eq. 2.10,

$$A_J = \frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}, \tag{2.10}$$

where $p_{T,1}$ and $p_{T,2}$ are the transverse momentum of the leading and the subleading jet respectively. Measure-

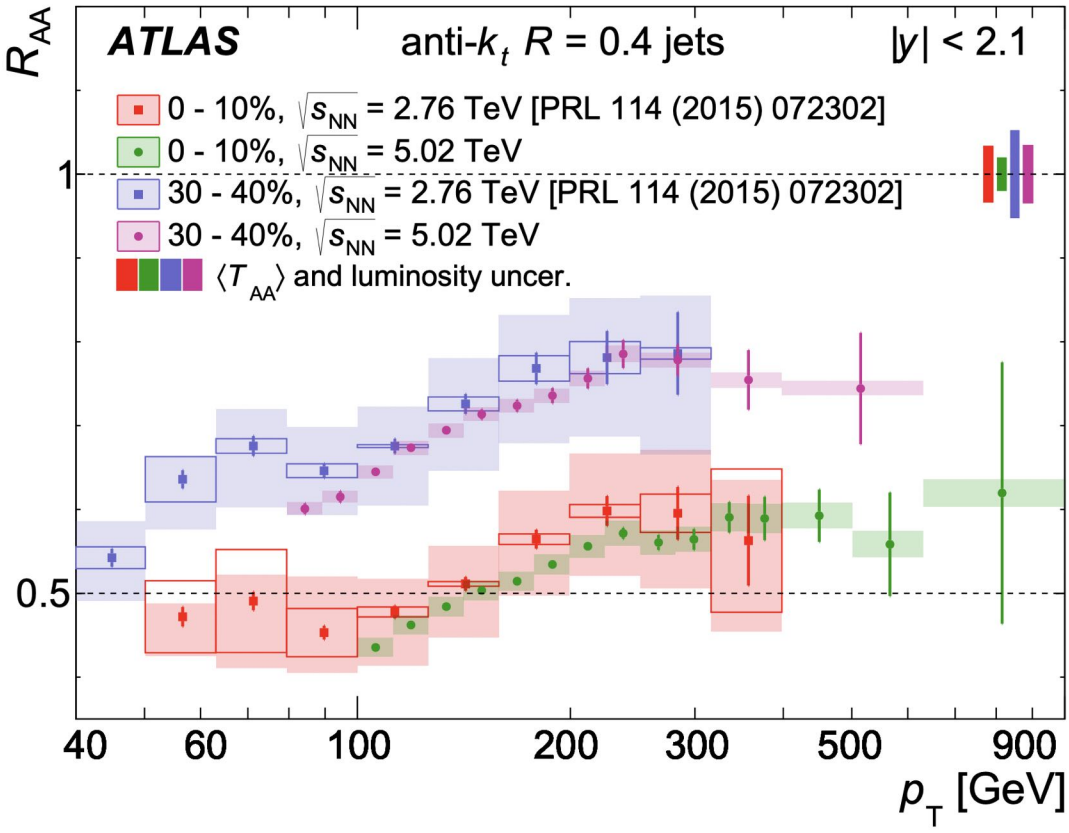

Figure 2.14: ATLAS measurement on the nuclear modification factor R_{AA} as a function of p_T at $\sqrt{s_{NN}} = 2.76\,\text{TeV}$. Figure from [46].

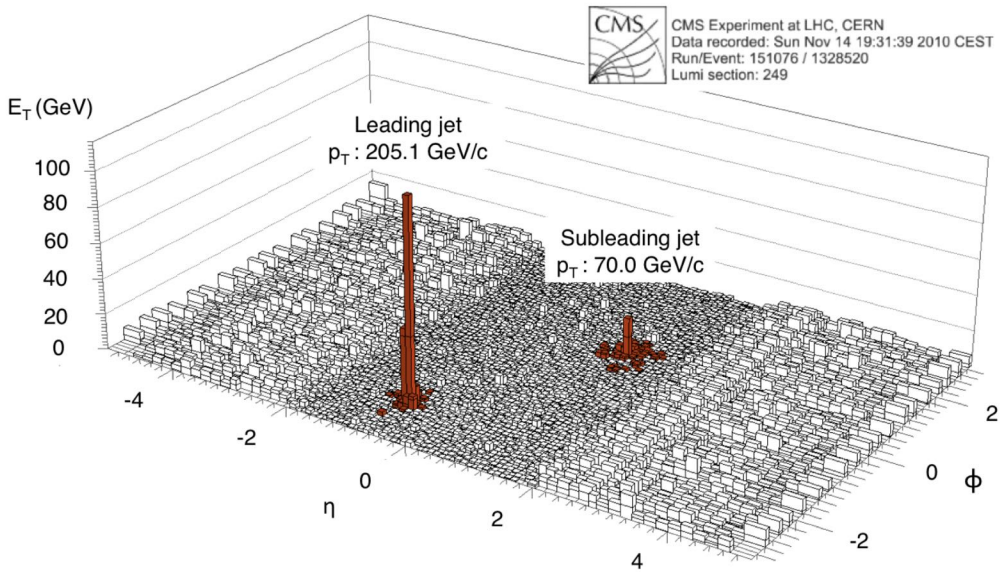

Figure 2.15: Dijet asymmetry in an PbPb collision recorded by the CMS experiment. Figure from [47].

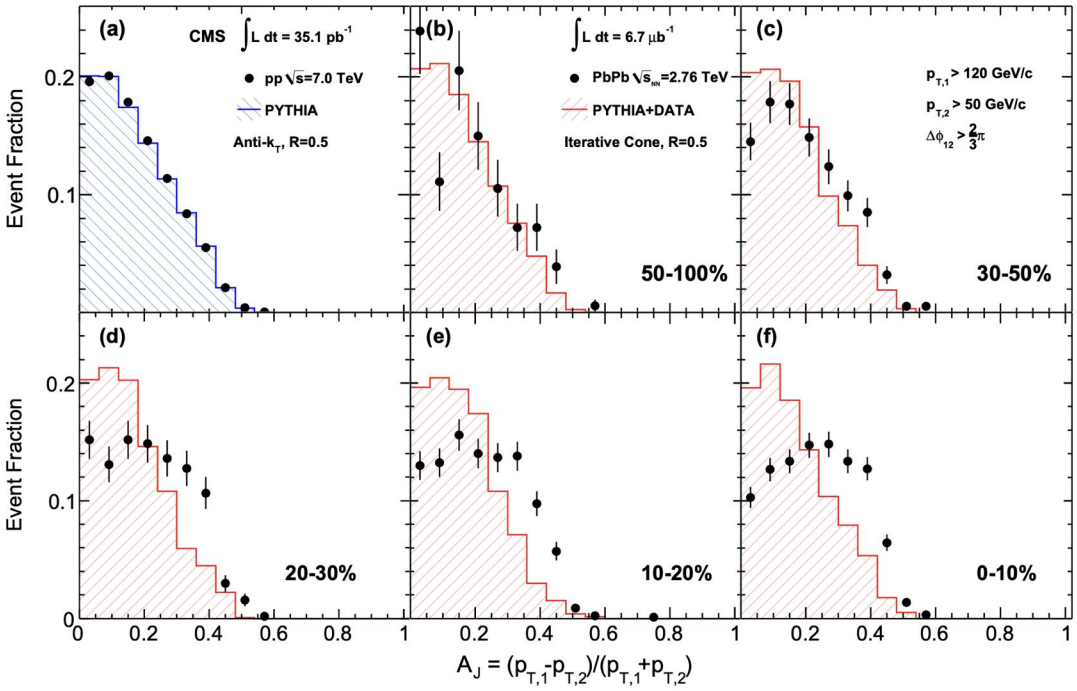

Figure 2.16: Distribution of the dijet asymmetry ratio (A_J), measured by the CMS experiment. Data are shown as black points, while histograms show (a) simulations with PYTHIA and (b)-(f) PYTHIA events embedded into PbPb data. Figure from [47].

ment of the distribution of A_J in PbPb collisions is reported by the CMS experiment [47]. For comparison, pp collisions simulated with the PYTHIA event generator are embedded into min-bias PbPb data. In this way, the embedded events are considered non-quenching references. Results (Fig. 2.16) can show a suppression at $A_J = 0$, especially in the most-central PbPb collisions (centrality $0-10\%$), indicating the dijets tend to be more unbalanced compared to those in peripheral collisions.

2.6.2.2 Fragmentation function and jet shape

The energy lost due to parton propagation in QGP is observed to be recovered by soft hadrons at large angles with respect to the jet axis. Observables such as the fragmentation function and jet shape help with understanding of the details about modifications due to the interactions with QGP. The jet fragmentation function aims to quantify the partition of the jet energy into its constituent particles, measured with variables

defined as,

$$z = \frac{p_{\parallel}^{track}}{p^{jet}},\ \xi = ln(\frac{1}{z}), \tag{2.11}$$

where $p_{\parallel}^{track}$ is the momentum component of the track along the jet axis. Low ξ values correspond to high p_T particles within the jet cone and vice versa. Measurements of the jet fragmentation functions in PbPb collisions of different centrality are shown in Fig. 2.17, as reported by the CMS experiment [48]. The distribution of ξ in peripheral PbPb collisions (centrality $70-100\%$) is in agreement with that measured in pp collision, showing no modification of the jet fragmentation function. However, for the most central collisions (centrality $0-10\%$), a significant enhancement on high ξ values ($p_T^{track} < 3$ GeV) is observed, with depletion in the intermediate region as compensation. The result has shown an enhancement of soft particle contribution to the jet energy and suppression of high-p_T particles in central PbPb collisions compared to pp collisions.

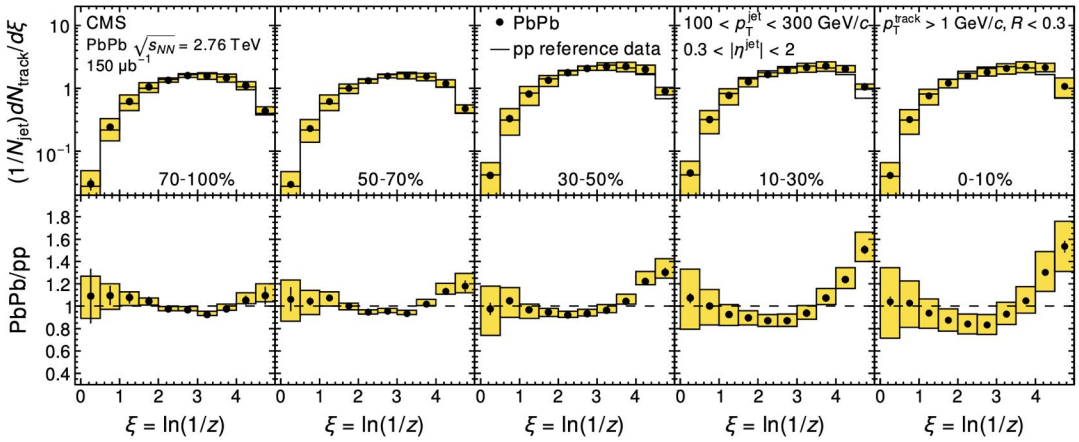

Figure 2.17: CMS measurements of the jet fragmentation functions in PbPb collisions (top) and their ratio to the pp reference (bottom), from peripheral (left) to central collisions (right). Figure from [48].

Further study of the jet energy redistribution and its dependence on distance from the jet axis in PbPb collisions can be done with measurements of the jet shape observable, which is defined as,

$$\rho(r) = \frac{1}{\delta r}\frac{1}{N_{jet}}\sum_{jets}\frac{\sum_{tracks\in[r_a,r_b)} p_T^{Track}}{p_T^{jet}}, \tag{2.12}$$

where the jet cone is divided into rings with radial width δr, and each ring has an inner radius $r_a = r - \delta r/2$ and an outer radius $r_b = r + \delta r/2$. Results from CMS measurements of the jet shapes [49] in PbPb collision

are shown in Fig. 2.18, with observation of the redistribution of jet energy to softer particles extending to large angles away from the jet axis, especially in the most central PbPb collisions.

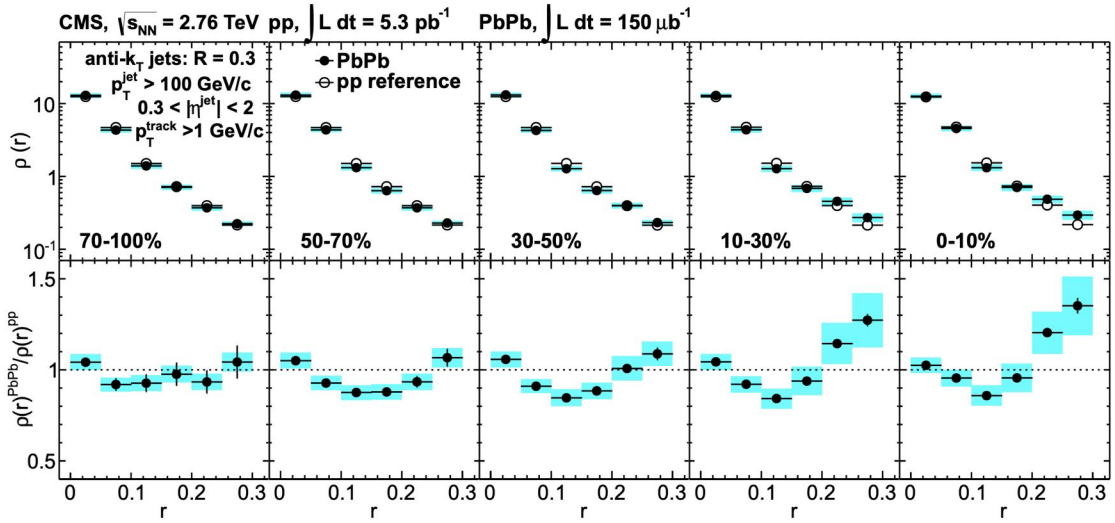

Figure 2.18: CMS measurement of the jet shapes in PbPb collisions (top) and their ratio to the pp reference (bottom), from peripheral (left) to central (right) collisions. Figure from [49].

2.6.2.3 Splitting function and jet substructure

The splitting function of QCD describes the $1 \to 2$ parton splitting process. Specifically, when two partons are seperated by an angle θ, the $1 \to 2$ splitting probability takes the form

$$dP_{i\to jk} = \frac{d\theta}{\theta} dz P_{i\to jk}(z), \tag{2.13}$$

where $P_{i\to jk}(z)$ are the Altarelli-Parisi QCD splitting functions [50] which depend on the momentum fraction z and the parton flavors i, j and k. Approximately, Eq. 2.13 for soft gluon emission from a hard gluon or quark can be described by

$$dP_{i\to ig} \simeq \frac{2\alpha_s C_i}{\pi} \frac{d\theta}{\theta} \frac{dz}{z}, \tag{2.14}$$

where α_s is the strong coupling constant and C_i is the Casimir factor (4/3 for quarks, 3 for gluons).

The QCD splitting function can not be directly measured due to the existence of a $\theta \to 0$ singularity.

However, efforts have been made [4], providing a semi-direct method to test the QCD splitting function by studying the 2-prong structure of jets. Specifically, the shared momentum ratio between the two prongs (z_g) can be measured in both pp and PbPb collisions, which takes the form

$$z_g = \frac{min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}}, \tag{2.15}$$

where $p_{T,1}$ and $p_{T,2}$ are the momenta of the two prongs and the g subscript denotes that z_g is obtained after a jet grooming algorithm is applied. Details of the jet grooming algorithm will be discussed in later sections. Analytic expression for Eq. 2.14 has been compared with z_g distribution measured in pp collisions with CMS Open Data, as shown in Fig. 2.19.

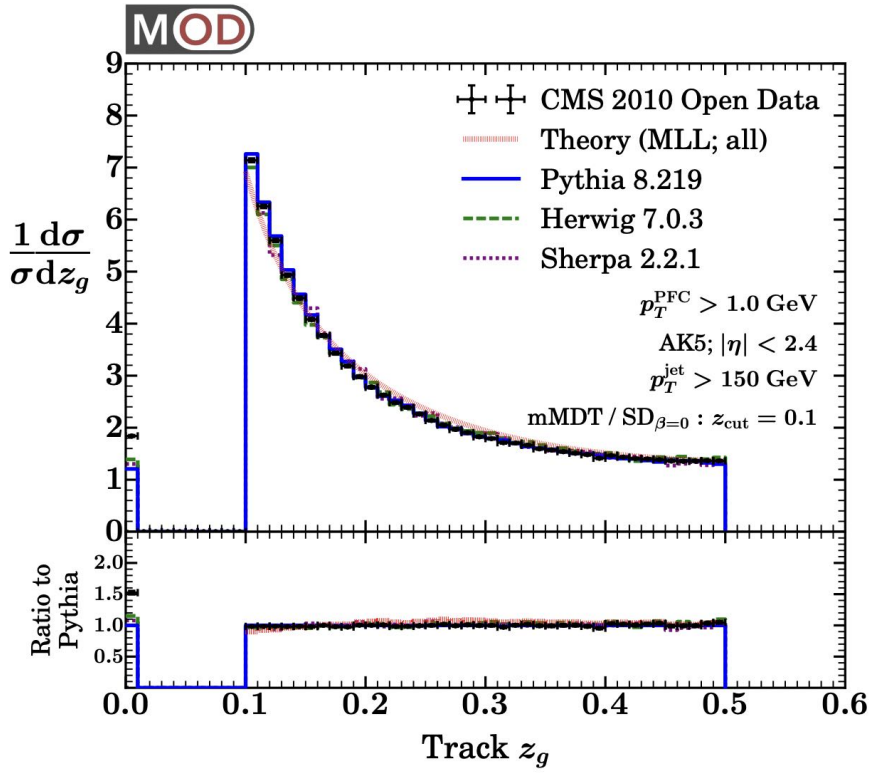

Figure 2.19: Distribution of z_g from CMS Open Data with analytic calculation. Figure from [4].

Such z_g distribution in PbPb collisions is also measured by the CMS experiment [51], leading to the observation of modification on jet substructure, shown in Fig. 2.20.

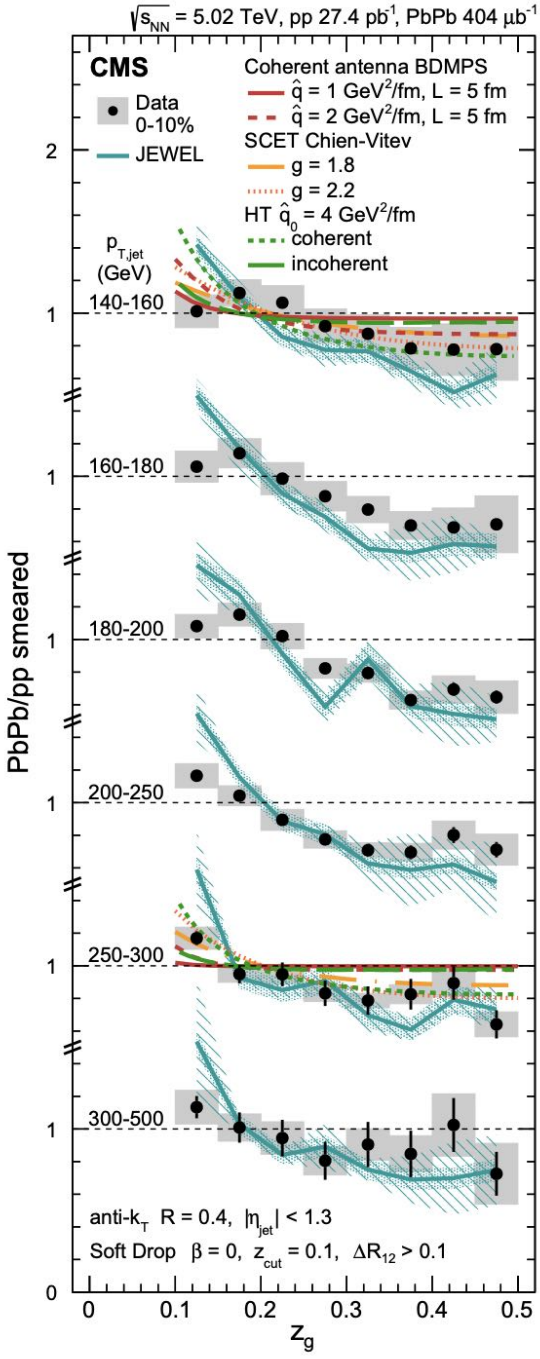

Figure 2.20: Distribution of z_g in PbPb collisions and smeared pp collisions, measured by the CMS experiment. Figure from [51].

CHAPTER 3

Monte Carlo Samples

Full Monte Carlo simulation involves two stages: (i) collision event generation and (ii) detector response. Physics processes are modeled by Monte Carlo event generators, which usually have a bunch of tunable parameters. Currently, various event generators are available on the market, with each having its own implantation of one or more processes. Detector system can also be modeled with Monte Carlo simulation framework GEANT4 [52], which is widely used in studies of high energy physics.

In this thesis, only the first part of the simulation is performed, without consideration of the detector response. In this section, the full procedure of how Monte Carlo samples are produced is introduced, including simulations on both the hard event generation and the soft thermal background. In addition, a thermal background subtraction method that is able to correct jet observables is discussed.

3.1 Dijet events

After hard scattering, scattered two partons will undergo the parton showering process individually, producing two back-to-back jets. Such kind of collision event is named as a dijet event simply because of the observation of two jets in one collision. Two Monte Carlo event generators are utilized to generate dijet events, which are introduced in this section.

3.1.1 PYTHIA

PYTHIA [53] is the most widely-used event generator in high energy physics and is able to produce pp collisions with various features, covering a wide range of energies. Currently, it is available with implementations of two programming languages, PYTHIA6 in Fortran and PYTHIA8 in C++. Both have the same underlying modeling of physics processes. In this thesis, PYTHIA8.2 is used. Among various processes of interest[1] that are implemented in PYTHIA8, only the hard QCD process is switched on in the simulation, which accounts for the $2 \rightarrow 2$ hard scattering process. In practice, the default tune 14 is used and a few parameters[2] are adjusted for biased selection on high p_T events, which are explained below.

- **PhaseSpace:pTHatMin** is the minimum average transverse momentum of scattered partons, with **100 GeV** used in this thesis.
- **PhaseSpace:bias2Selection** is the switch of biased sampling from the phase space. The default value is off while **On** is used.
- **PhaseSpace:bias2SelectionRef** determines that unit event weight should be assigned to events that are sampled with $\hat{p}_{T,Ref}$ from the phase space, if above parameter is On. The value of **100 GeV** is used.
- **PhaseSpace:bias2SelectionPow** determines the oversampling at $\hat{p}_T$ should be done with an amount of $(\hat{p}_T / \hat{p}_{T,Ref})^{Pow}$. The value of **4** is used.

3.1.2 JEWEL

JEWEL [54] is a Monte Carlo event generator that is able to simulate jet evolution in heavy ion collisions, which implements a fully microscopic description of jet evolution in medium, including coherence effects. In JEWEL, all partons coming from the initial hard scattering process undergo collisions with medium partons

[1] Details can be found: https://pythia.org/manuals/pythia8245/ProcessSelection.html
[2] Details can be found: https://pythia.org/manuals/pythia8245/PhaseSpaceCuts.html

and suffer both radiative and elastic energy loss. As a result, recoiling medium partons carry away energy and momentum from the jet. Considering the fact that medium partons are heated up by a jet and eventually become part of the jet, such a medium response would definitely introduce modifications to the jet substructure.

In practice, JEWEL can be used in the study of: (i) the showering process of the initial hard scattered parton in a medium and (ii) medium response. Note that even though JEWEL claims to be able to simulate jet evolution in heavy ion collisions, it does not have a description of the thermal background, which comes from soft rather than hard processes. In this thesis, such thermal background is named "uncorrelated", as it has nothing to do with the hard scattering process. While the "correlated" background refers to the medium response, consisting of recoiling thermal partons due to interactions with the hard scattered partons.

Treatment of the recoiling partons is tricky. It is easy to switch on or off the medium response effects in simulation, while it is rather difficult to do so in real experiments. With JEWEL, events can be generated with or without a medium response. When running without medium response ("recoil off"), recoiling partons do not show up in the event. Thus, inclusive or inter-jet observables can be compared to (background-subtracted) experimental data. On the contrary, when running with a medium response ("recoil On"), information about recoiling partons is stored in the event, which makes it possible in the studies of the modification introduced by the medium response. Such recoiling partons can not be subtracted with experimental background subtraction techniques. As a result, a special subtraction algorithm ("4MomSub") is introduced with jet substructure observables studied [55]. A few parameters[1] are configured as described below.

- **PTMIN** is the minimum average transverse momentum, with **100 GeV** in use.
- **MASS** is the mass number of colliding nucleus (**208.** for Pb).
- **SQRTS** is the center-of-mass energy ($\sqrt{s_{NN}}$) of the collising system, with **5020. GeV** in use.

[1] Details can be found from the user manual [54].

- **PROCESS** is the selected hard process, with **"PPJJ"** standing for hard scattering (dijet).
- **KEEPRECOILS** controls whether or not recoils effects should be included (**T** for true).
- **WRITESCATCEN** controls whether or not information about thermal momentum component should be recorded (**T** for true).
- **WRITEDUMMIES** controls whether or not dummy particles (used for 4MomSub) should be recorded (**T** for true).

Parameters related to medium properties are also configurable and are described below.

- **TAUI** is the initial time τ_i (fm/c), with **0.6** in use.
- **TI** is the initial temperature T_i (GeV), with **0.4** in use.
- **CENTRMIN** is the lower end of centrality range (%) to be simulated, with **0.** in use.
- **CENTRMAX** is the upper end of centrality range (%) to be simulated, with **10.** in use.

In this thesis, JEWEL events are generated with the medium response ("recoil On"), while the "4MomSub" is not performed. Dijet events generated from PYTHIA and JEWEL are mixed with uncorrelated thermal background, after which an experimental background subtraction method is applied. Details about the generation of such thermal background as well as the mixing and subtraction techniques are described in later sections.

3.2 Uncorrelated underlying events

The uncorrelated underlying events, as described briefly in the previous section, consist of particles that do not originate from the hard processes, but rather - from soft interactions. Such events are often referred to

(i) the pileup events in pp collisions and (ii) the underlying events in heavy ion collisions, which both are the results of simultaneous proton-proton collisions in a single beam bunch crossing. The high-luminosity LHC is expected to deliver an average pileup of 200 pp collisions per bunch crossing, while the underlying event in the most central heavy ion collisions at LHC energy may lead to a background p_T of more than 300 GeV per unit area in $\eta - \phi$ space (Fig. 3.1).

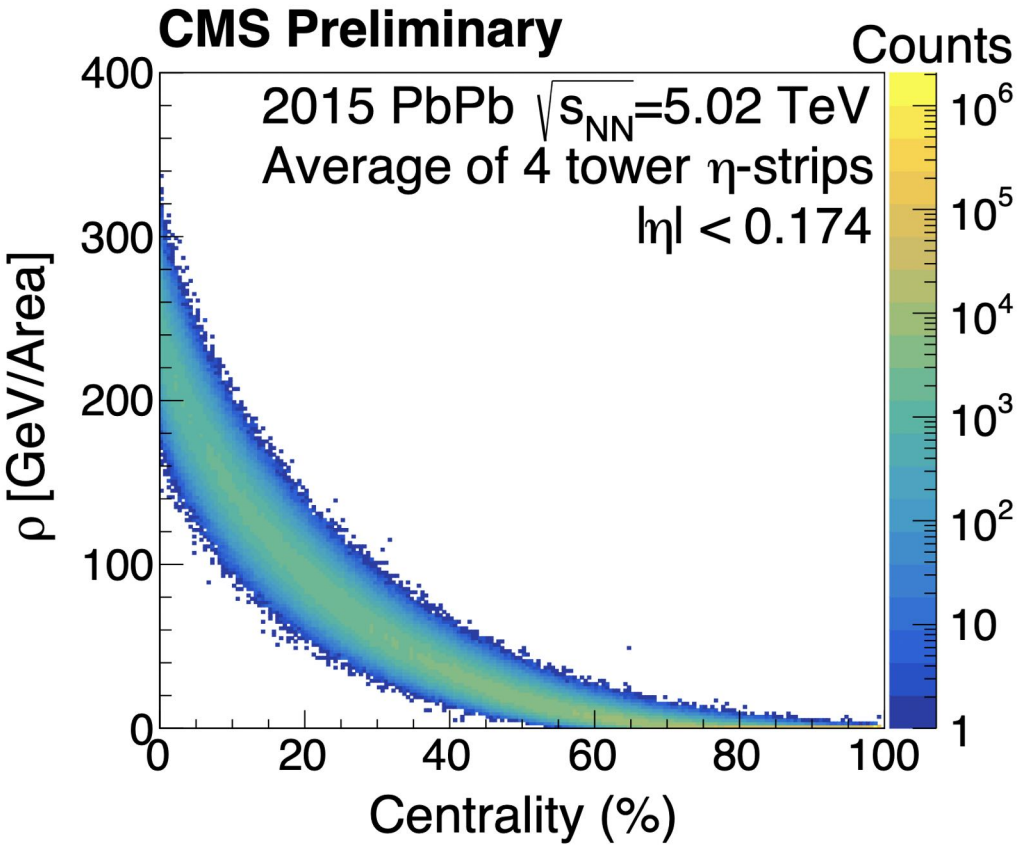

Figure 3.1: The energy density deposited in the calorimeter towers of the CMS detectors in PbPb collisions as a function of the event centrality. Figure from [56].

A data-driven approach to the generation of a thermal background is applied. The uncorrelated underlying event is generated following a Boltzmann distribution for the transverse momentum of the particles, which are distributed uniformly in pseudorapidity η and azimuthal angle ϕ (Fig. 3.2 and Fig. 3.3). This method was previously used in ref. [57]. Uncorrelated underlying events are generated with configurations of two different multiplicity settings, corresponding roughly to the mid-central (40-50%) and the most-central (10-

20%) PbPb collisions, inferred from published experiment results [58]. In this thesis, the two multiplicities are distinguished as a low multiplicity with a total multiplicity of 1700 and a high multiplicity with a total multiplicity of 7000. Details about the two multiplicity configurations are listed in Tab. 3.1.

Values	Mid-central	Most-central
Pseudorapidity Interval	$\|\eta\| < 3$	$\|\eta\| < 3$
Total Event Multiplicity	1700	7000
Corresponding Centrality Interval	40-50%	10-20%
Average Yield $\langle \frac{dN}{d\eta} \rangle$	283	1167
Average Transverse Momentum $\langle p_T \rangle$	0.9 GeV	1.2 GeV
Average Transverse Momentum Density $\langle \rho \rangle$	35.0 GeV/Area	227.1 GeV/Area

Table 3.1: Description of the simulated underlying events.

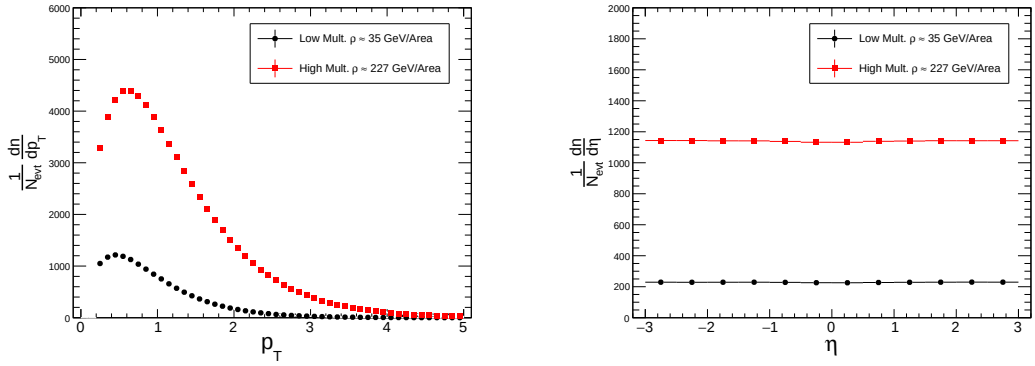

Figure 3.2: p_T spectrum (left) and yield (right), for particles from uncorrelated underlying events generated with a data-driven approach at two different multiplicities.

3.3 Background Subtraction

In this thesis, a single embedded event is made of a dijet event and an uncorrelated underlying event, which respectively intimate the hard scattering and the soft thermal background in heavy ion collisions. An embedded PYTHIA event does not have any quenching effect, even though a thermal background is mixed in, thus making it a control set. In contrast, the embedded JEWEL event has a quenching effect and thermal background, as is expected in real heavy ion collisions.

The underlying event in PbPb collisions is observed to fluctuate per event, with its p_T deposit being a

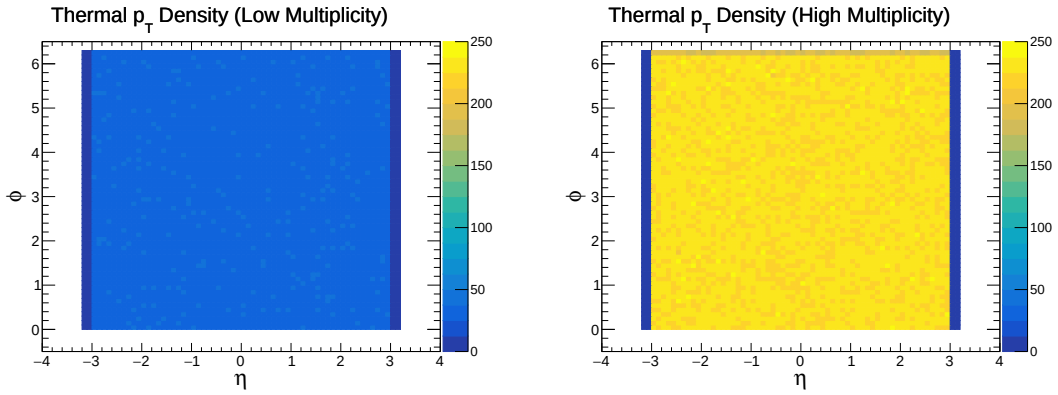

Figure 3.3: Average p_T deposit in $\eta - \phi$ plane, for particles from uncorrelated underlying events generated with a data-driven approach at low multiplicity (left) and high multiplicity (right).

function of η and ϕ. Due to intense pileup events in pp collisions and underlying events in PbPb collisions, efficient techniques are crucial in order to mitigate the impact of large backgrounds on jet kinematics and jet substructure. Luckily, various background subtraction methods are available.

One category of background subtraction methods depends on a specific jet definition. For example, Area Subtraction [59], an area-based subtraction method that is able to correct jet's 4-momenta from its area and an estimated energy density. The area, a jet property, is unique to the jet definition. While another category of subtraction methods try to correct particles prior to jet clustering, such as Constituent Subtraction [60], Iterative Constituent Subtraction [61], SoftKiller [62], PUPPI [63] and jet cleansing [64].

3.4 Constituent Subtraction

In this thesis, the Constituent Subtraction (CS) algorithm is used for background subtraction on embedded events. Three implementations, jet-by-jet, event-wide and iterative CS are in the algorithm family, sharing a similar design on background density (ρ) estimation and on ghost subtraction.

The background density ρ, as its name suggested, is the amount of p_T from the underlying event per unit

area on the $\eta - \phi$ plane. An event is divided into patches on the $\eta - \phi$ plane (Fig. 3.4), with each has

$$p_{T,patch} = \sum_{i \in patch} p_{T,i}. \tag{3.1}$$

Thus the estimated ρ can be calculated as

$$\rho = median_{patches}\{\frac{p_{T,patch}}{Area_{patch}}\}, \tag{3.2}$$

where the median instead of the average is used to avoid over-estimation due to high-p_T patches. In the following subsections, how the background density ρ is used in three subtraction methods will be introduced. These three methods have very little difference in their implementations, while they share the same idea of how to use the background density ρ. Simply speaking, "ghosts" carrying negative momenta proportional to ρ are added to events, which behave like momentum consumers. The way those ghosts consume momentum is exactly the same as subtraction, which will be introduced in the following subsections[1].

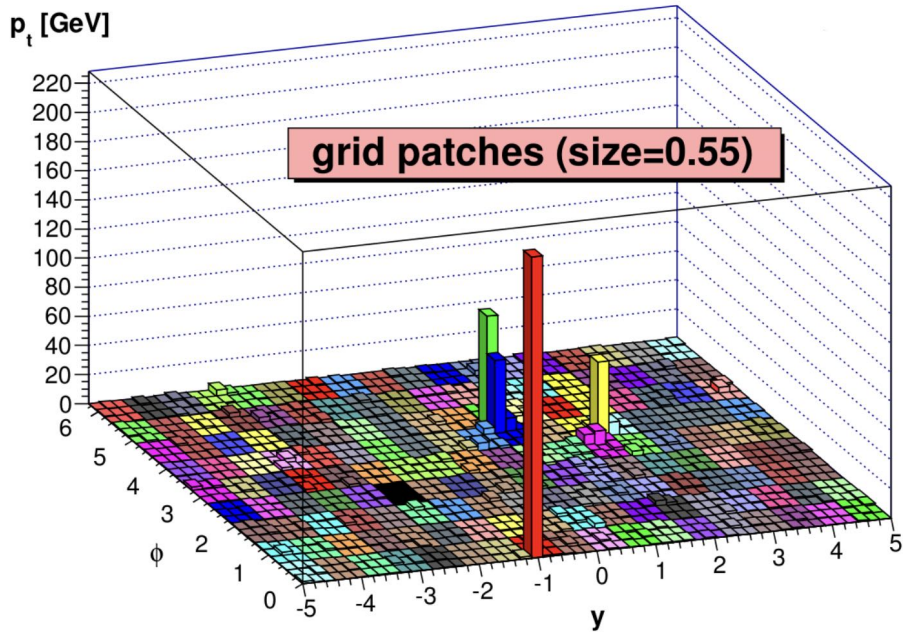

Figure 3.4: Patches for estimation on background density ρ[2].

[1] Materials from https://indico.cern.ch/event/649482/contributions/2993293/attachments/1687676/2714424/PeterBerta_CS_17.7.2018.pdf

[2] Figure from Boost2012 https://indico.cern.ch/event/179612/timetable/?view=standard

3.4.1 Jet-by-jet

The subtraction procedure starts with adding uniformly distributed ghosts to the event, of which each populates a small area (implemented as a free parameter A_{ghost}) on the $\eta - \phi$ plane and carries a negligible amount of momentum. In the second step, particles and ghosts are clustered into jets with jet-finding algorithms. The momenta that ghosts carry are so few; thus, they can be safely clustered into jets without introducing significant differences in the jet kinematics. In the third step, the ghosts that are clustered into jets are set to carry negative momenta proportional to the estimated background density ρ. In the last step, the ghosts are paired with constituents based on a specific metric. The 4-momenta of each constituent is corrected by subtracting the 4-momenta of its paired ghost. The full procedure is illustrated in Fig. 3.5.

The metric of how a ghost is paired with a jet constituent is determined based on their "distance", which is described by

$$\Delta R_{i,k} = p_{T,i}^{\alpha} \cdot \sqrt{(y_i - y_k^g)^2 + (\phi_i - \phi_k^g)^2}, \tag{3.3}$$

where i and k are the indices of the jet constituent and the ghost respectively, α is a free parameter and the subscript g indicates a quantity that is associated with the ghost. The introduction of a positive α indicates that a jet constituent with lower p_T is more likely to be paired to a ghost. In other words, jet constituents with lower p_T (being mostly particles from thermal backgrounds) are subject to correction in the subtraction procedure.

The correction starts from the ghost-particle pair with the lowest distance and a jet constituent's momen-

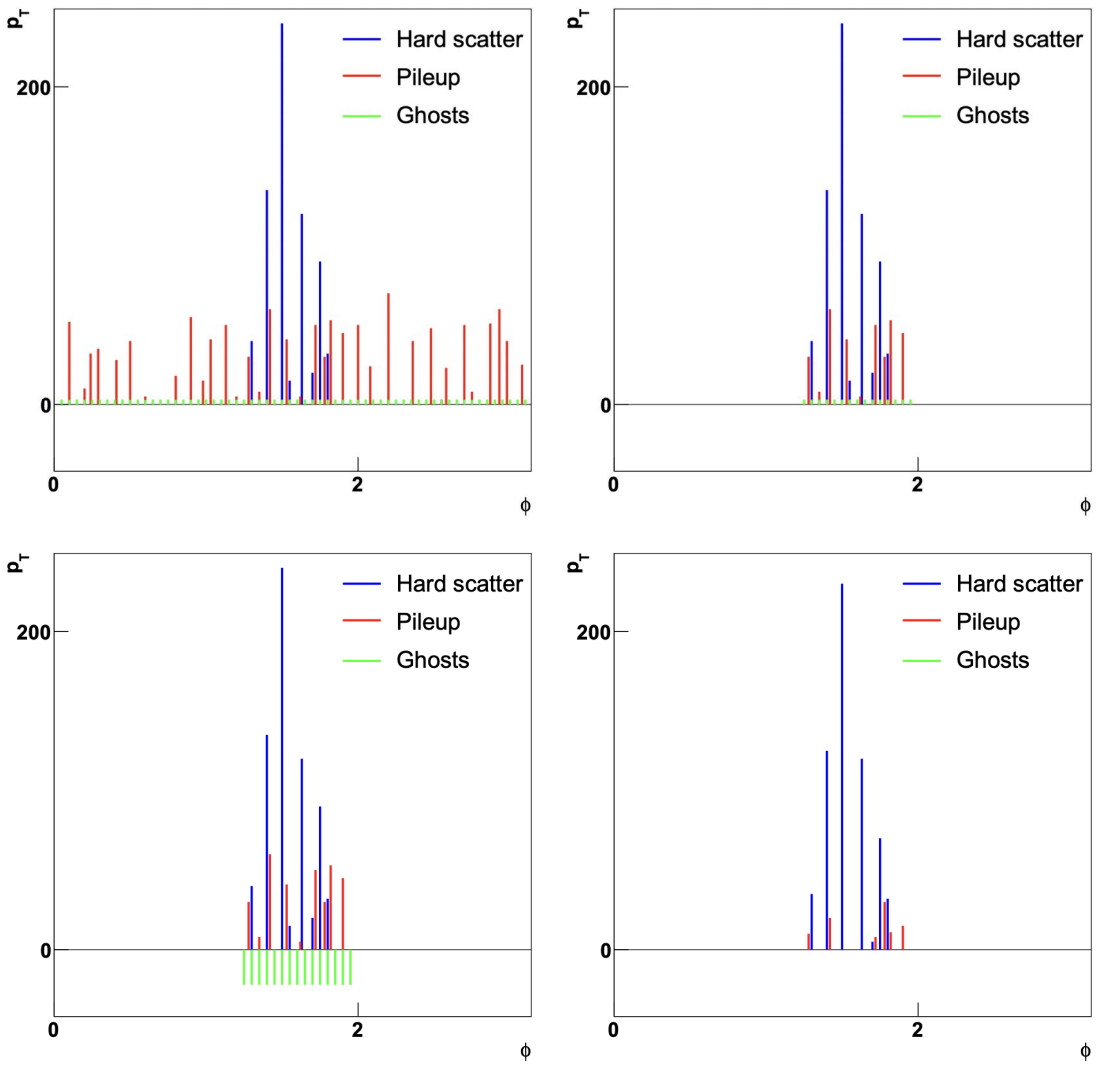

Figure 3.5: Procedure of the jet-by-jet constituent subtraction algorithm with 4 steps: (i) adding ghosts to an event (upper left), (ii) clustering particles and ghosts into jets (upper right), (iii) setting ghosts' momenta to be negative and proportional to ρ (lower left), (iv) subtracting ghosts' momenta from their paired constituents (lower right).

tum is corrected by following

$$
\begin{aligned}
&if\ p_{T,i} \geq p^g_{T,k}: \\
&\quad p_{T,i} \to p_{T,i} - p^g_{T,k}, \\
&\quad p^g_{T,k} \to 0, \\
&otherwise: \\
&\quad p_{T,i} \to 0, \\
&\quad p^g_{T,k} \to p^g_{T,k} - p_{T,i}.
\end{aligned}
\tag{3.4}
$$

Such subtraction procedure stops if

$$\Delta R_{i,k} > \Delta R_{max}, \tag{3.5}$$

where ΔR_{max} is a free parameter.

3.4.2 Event-wide

The event-wide constituent subtraction implements a procedure that is very similar to the jet-by-jet method. It starts by adding ghosts to the event. However, ghosts are not clustered into jets, which is the case in the jet-by-jet method. So the name "event-wide" comes due to the fact that the ghosts and particles are paired event-wide, with the same metric in the distance as Eq. 3.3. Then the subtraction begins on the (event-wide) ghost-particle pair with the lowest distance, and the same subtraction method as Eq. 3.4 is used. The full procedure of the event-wide constituent subtraction is shown in Fig. 3.6.

In this thesis, the event-wide constituent subtraction method is used with free parameters configured as

$$
\begin{aligned}
&A_{ghost} = 0.005, \\
&\alpha = 1, \\
&\Delta R_{max} = 0.3.
\end{aligned}
\tag{3.6}
$$

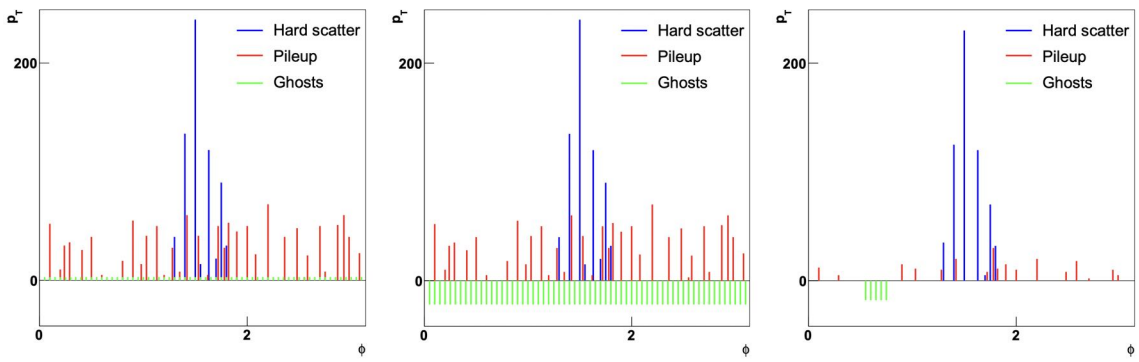

Figure 3.6: Procedure of the event-wide constituent subtraction algorithm with 3 steps: (i) adding ghosts to an event (left), (ii) setting ghosts' momenta to be negative and proportional to ρ (mid), (iii) event-wide pairing and subtracting ghosts' momenta from their paired constituents (right).

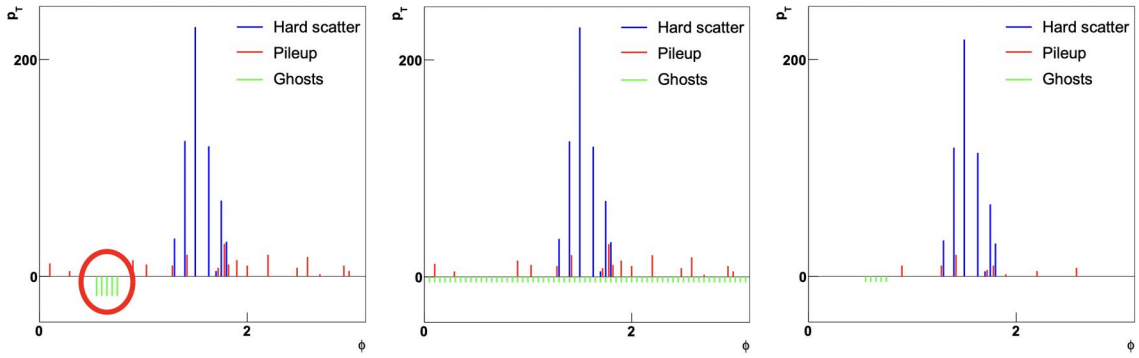

Figure 3.7: Procedure of the iterative event-wide constituent subtraction algorithm: (i) leftover ghosts after the first iteration of the event-wide constituent subtraction (left), (ii) setting ghosts' momenta to be negative and proportional to ρ after redistribution (mid), (iii) a second iteration of event-wide pairing and subtraction (right).

Comparison between different subtraction methods and between different free parameter sets will be introduced in later sections.

3.4.3 Iterative

In the event-wide constituent subtraction method, a local cold spot in the thermal background produces leftover ghosts carrying momenta that are not subtracted, which can be seen from Fig. 3.6 (right). Thus an improvement, the iterative event-wide constituent subtraction method, is introduced. After the first iteration is applied, which is the same as the event-wide subtraction, leftover ghosts are redistributed uniformly on the

$\eta - \phi$ plane. And a second iteration takes action to handle those leftover ghosts, as shown in Fig. 3.7. The second iteration may utilize different free parameter set (A_{ghost}, α and ΔR_{max}). More iterations other than two could also be applied.

CHAPTER 4

Jet Reconstruction

A jet is a topological object reconstructed from collimated entities as a proxy collection of the sprayed hadrons originating from the fragmentation of a highly energetic parton. In experiment and also in theory, the definition of a jet, including its energy and its boundary (or area), is ambiguous. Firstly, the definition of a jet depends on its clustering algorithm, which determines the composition (constituents) of a jet and its boundary. Secondly, a jet may be clustered from full information at the hadron level, or partial information which contains only energy deposits in detectors or reconstructed tracks. In practice, jet algorithms are required to be simple to implement in experiment analysis and theoretically calculable in theory. In addition, two more properties are desired:

- Invariant under the radiation of infinitely soft particles,
- Invariant under collinear splittings into hard fragments.

The above two requirements are often referred to as the infrared safety and the collinear safety (or jointly the IRC-safety), as illustrated in Fig. 4.1 and Fig. 4.2 from Ref. [65]. Any disobey of these two requirements in the design of a jet clustering algorithm will lead to biased jet spectra and divergence in theoretical calculations.

In this thesis, full information of simulated events is accessible at the hadron level, without detector response. A jet clustering algorithm is performed on events, of which background is subtracted with the event-wide constituent subtraction method. In this chapter, the family of jet clustering algorithms will be introduced, together with the jet grooming techniques. The combination of jet clustering and jet grooming helps access the jet substructure, which is the topological structure of a jet containing information about the

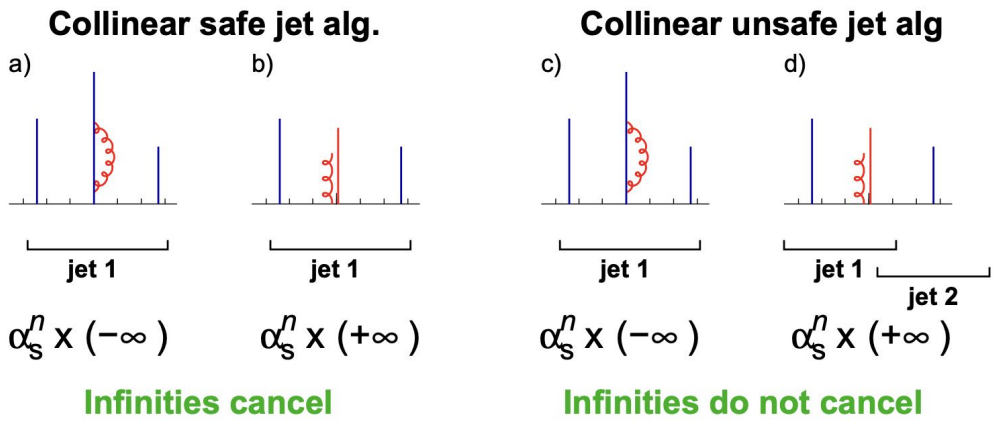

Figure 4.1: Illustration of collinear safety (left) and collinear unsafety (right).

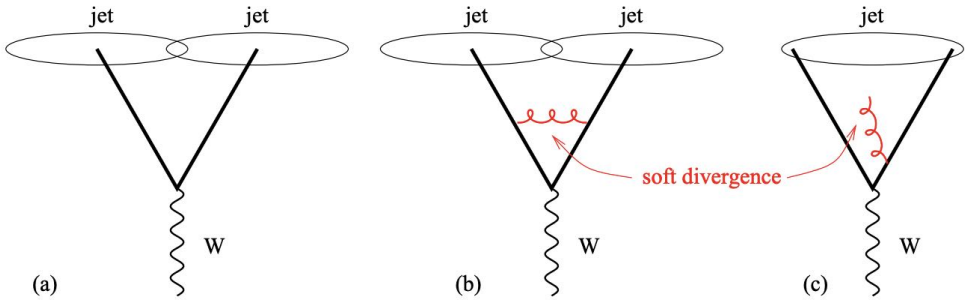

Figure 4.2: Illustration of collinear safety (a) and collinear unsafety (b,c). The addition of a soft gluon converts the event from having two jets (b) to just one jet (c).

jet showering process.

4.1 Jet clustering algorithms

Currently, in high energy physics, a family of jet clustering algorithms known as sequential recombination algorithms are employed, which iteratively pair entities (particles or pseudojets[1]) to reconstruct jets. Such clustering procedure, also referred to as jet finding, terminates until particles are exhausted or a specific terminating condition is reached.

In the clustering, a distance $d_{i,j}$ between entities i and j, and a distance d_{iB} between entity i and the beam

[1] Here, it refers to the intermediate results of reconstructed jets.

are defined as,

$$
\begin{aligned}
d_{ij} &= min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \\
d_{iB} &= k_{t,i}^{2p},
\end{aligned}
\tag{4.1}
$$

where $\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ and $k_{t,i}$, y_i and ϕ_i are transverse momentum, rapidity and azimuthal angle of entity i. In addition to the radius parameter R, a parameter p is introduced to govern the relative power of the transverse momentum versus angular distance Δ_{ij}. The clustering algorithm iteratively identifies the smallest distance, and if it is d_{ij} then the two entities are combined, while if it is d_{iB} then entity i is labeled as a jet and is removed from entity list.

Different values of parameter p correspond to different clustering algorithms [66], with $p = -1$ referring to the anti-k_t algorithm, $p = 1$ referring to the k_t algorithm and $p = 0$ referring to the Cambridge/Aachen (C/A) algorithm. Though sharing similar metrics in measuring distances, these algorithms behave differently. The k_t algorithm tends to pair soft particles first. On the contrary, the anti-k_T algorithm is more likely to pair hard particles first. Since $p = 0$ in the Cambridge/Aachen algorithm, the metric no longer depends on the transverse momentum. Thus the C/A algorithm tends to pair particles with smaller angular distances first. In addition to differences in the ordering, differences in the boundaries of reconstructed jets are there. The anti-k_T is commonly used as the default jet finding algorithm due to its constrain on the jet area, producing jets with circular areas. For illustration, the behaviors of these algorithms are shown in Fig. 4.3.

All the sequential clustering algorithms described above are IRC-safe and are implemented in the FastJet framework [67].

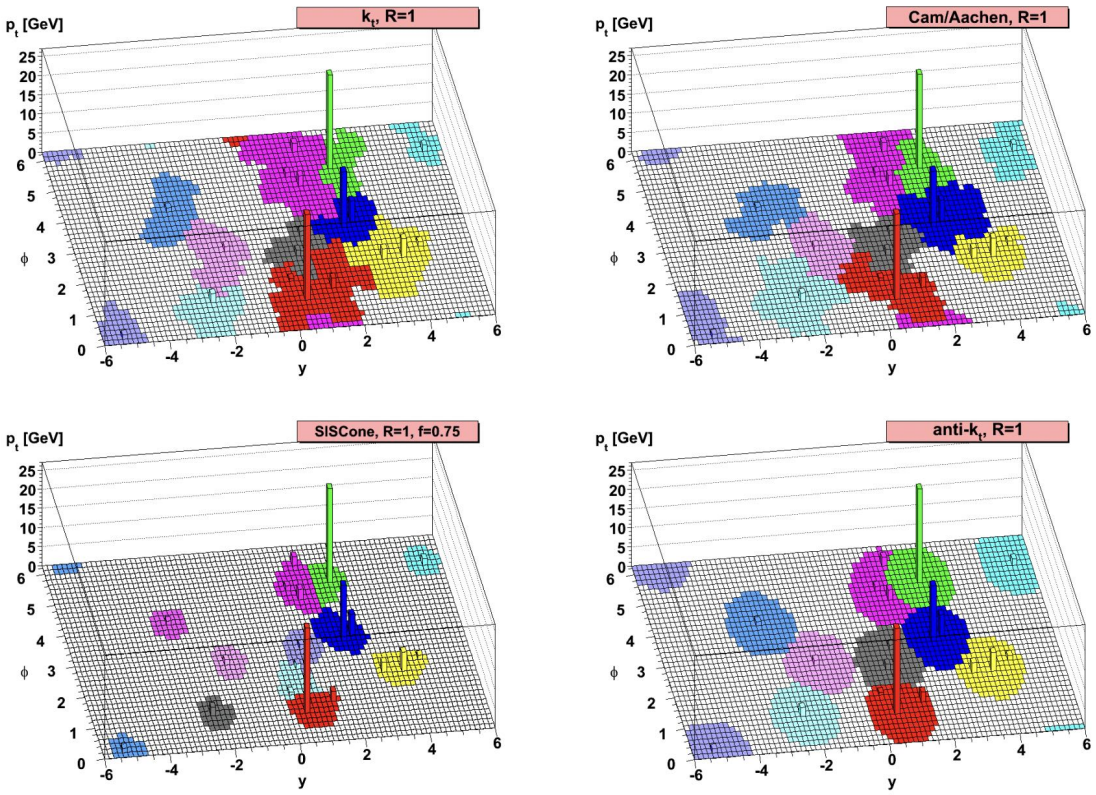

Figure 4.3: Illustration of jet clustering algorithms. Figure from [66].

4.2 Jet grooming

Jet grooming techniques are mostly used in the context of boosted jets and jet substructure, with improvements in the reconstruction precision and benefits in mitigation of pileup and underlying events. Compared to background subtraction methods, jet grooming techniques are not generic subtraction methods. The most frequently used grooming methods are filtering [68], trimming [69], pruning [70] and soft drop [71], each of which has free parameters that control the performance of configured groomers. Groomers based on machine learning [72] are also implemented and tested recently. These methods often work with subjets that are determined by the Cambridge/Aachen (C/A) algorithm applied to an anti-k_t jet. Specific algorithms or criteria related to grooming are then applied to these subjets to eliminate soft and/or wide-angle contributions to a jet, which are the most likely to be contaminated by thermal backgrounds.

The soft drop grooming starts from jets defined with the anti-k_t algorithm. Constituents of an anti-k_t jet with a radius R_0 are reclustered with the Cambridge-Aachen algorithm to form a pairwise clustering tree with an angular-ordered structure (Fig. 4.4). Such declustering takes advantage of the fact that the C/A algorithm tends to combine entities that are close to each other, prior to entities that are far away.

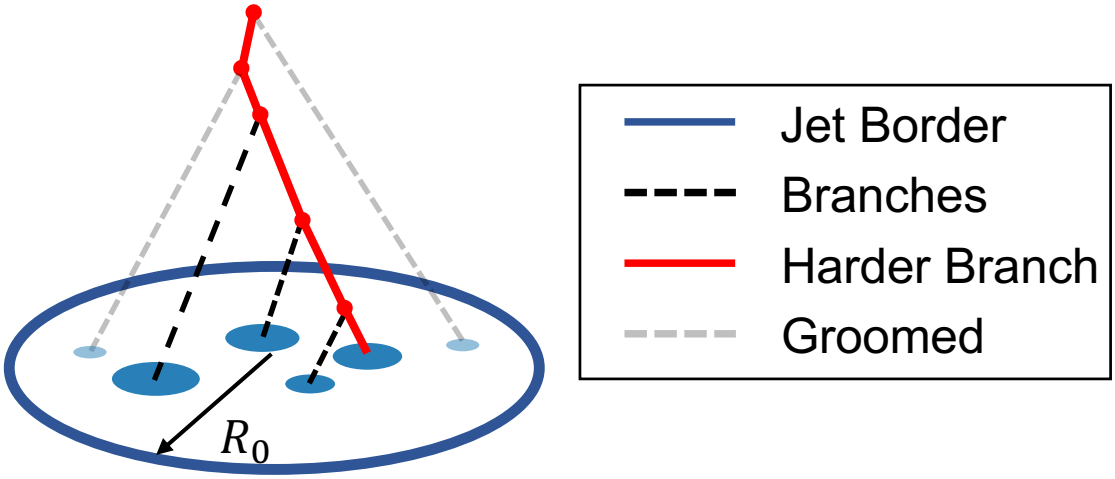

Figure 4.4: Pairwise clustering tree with an angular-ordered structure, of which soft branches are groomed by a soft drop groomer.

The soft drop declustering procedure has two free parameters, a threshold of the shared momentum ratio

z_{cut} and an angular power β, and is implemented as:

1. Undo the last step of C/A clustering and break the jet j into two separated subjets as j_1 and j_2.

2. Given the soft drop condition,

$$z \equiv \frac{min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut}(\frac{\Delta R_{12}}{R_0})^{\beta}, \tag{4.2}$$

   where z is the momentum fraction carried by the softer subjet, $\Delta R_{12} = \sqrt{(y_1 - y_2)^2 + (\phi_1 - \phi_2)^2}$ is the angular distance between subjets and R_0 the jet radius. If the two seperated subjets pass the soft drop condition, then the declustering stops and the jet j is the final soft-drop jet.

3. Otherwise, The subjet with smaller p_T is groomed. Redefine j to be the subjet with larger p_T and repeat the soft drop declustering procedure.

Such procedure is illustrated in Fig. 4.5.

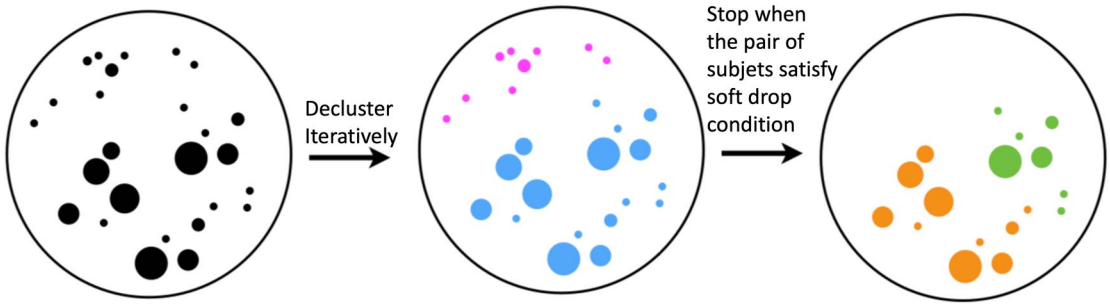

Figure 4.5: Illustration of the soft drop grooming.

4.3 Groomed substructure observables

The declustering procedure of the soft drop method is able to expose the topological structure of a jet and group its constituents into a hierarchical tree. Soft branches that can't pass the grooming criteria, corresponding typically to the soft and wide-angle radiation, are discarded. In contrast, the two subjects that can pass

the criteria carry information about the hard splitting that happens in the early stage of the parton showering. Such information can be interpreted and is useful in the study of the jet quenching phenomenon.

The first splitting that satisfies the soft drop condition (Eq. 4.2), the "2-prong" structure (Fig. 4.6), provides access to the branching kinematics. In practice, groomed substructure observables can be measured, for

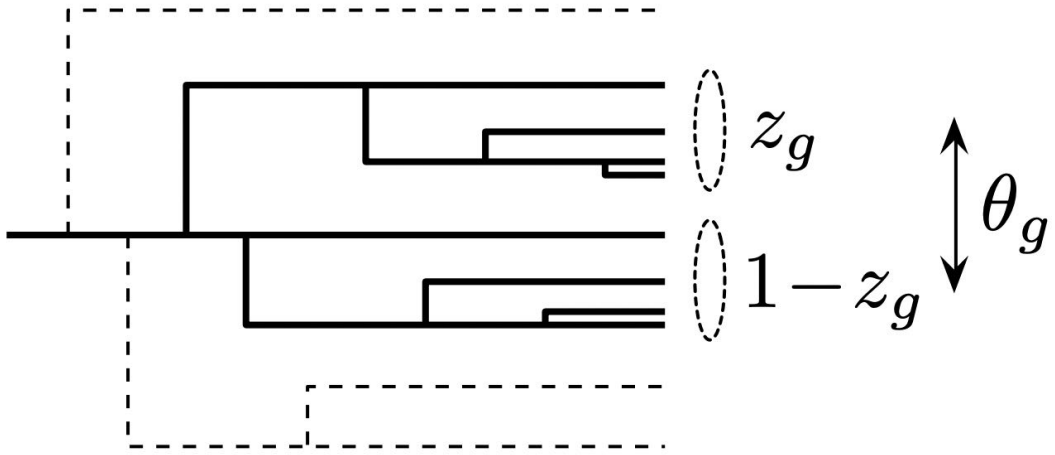

Figure 4.6: Illustration of the "2-prong" structure of a jet. Deashed lines correspond to subjets that are groomed by the soft drop groomer. θ_g is defined as R_g/R_0. Figure from [4].

example, z_g, which is the momentum fraction carried by the relatively softer subjet. The subscript g denotes such observable is obtained from a groomed jet. Other observables include the angular distance between the two separated subjets R_g and groomed jet mass m_g.

These groomed substructure observables are sensitive to thermal backgrounds and jet quenching, which will be analyzed in later sections.

4.3.1 Sensitivity to thermal backgrounds

Measurements on the groomed substructure observables have opened a new front in heavy ion jet physics. However, background effects are not well understood. Local background hot spots in heavy ion collisions can lead to incorrect identification of the splitting that is unrelated to the jet showering process. Though such

background effects can be avoided in the background-free region of phase space, namely at sufficiently large p_T and/or smaller jet radius, the understanding of how the background subtraction affects jet substructure and associated groomed substructure observable is imperative.

4.3.1.1 Mistagging of the subleading prong

In this section, studies on the mistagging of the subleading prong in the presence of thermal backgrounds are discussed, based on Ref. [73].

As the soft drop grooming method can be utilized to reveal the "2-prong" structure of a jet, a question arises as whether the thermal background affects tagging on the leading and subleading prongs, with the following cases that can happen:

- Both the leading and subleading prongs are correctly reconstructed and tagged.
- Both prongs and correctly identified, but swapped. In this case, z_g and R_g are not affected.
- The leading prong is correctly tagged, while the subleading prong is reconstructed from background hot spot (mistag), as shown in Fig. 4.7.
- The subleading prong is groomed away.
- the subleading prong is outside the jet, while some of its constituents may be present.
- Other cases, for example several above cases all happen at the same time.

Simulations have been done by embedding PYTHIA jets into a thermal background that is comparable to that in the most central heavy ion collisions. Constituent subtraction is applied to the embedded event to mitigate the impact introduced by thermal backgrounds. The performance of locating the splitting of interest is studied across various groomers, with the measurement of the subleading prong tagging purity, as shown

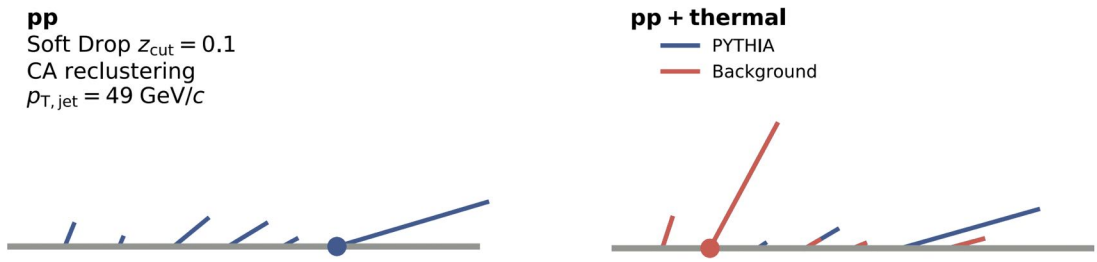

Figure 4.7: Example of mistagging of the subleading prong in embedded events. In the case of thermal background (right), a background hot spot at large angle passing the grooming conditions leading to the true subleading prong being aborbed into the leading prong, thus a mistagging happens.

in Fig. 4.8. These studies have shown that the mistagging rate is reduced with jet p_T over 200 GeV, for soft drop groomers with $z_{cut} = 0.1$ and $\beta = 0$. The soft drop groomer with such a free parameter set is used in later studies.

4.3.1.2 Dependence on background subtraction

In this section, the performance of the constituent subtraction method is studied across its three implementations, which are the "jet-by-jet", the "event-wide" and the "iterative". Descriptions about the procedure of each implementation is discussed in Sec. 3.4. The soft drop grooming method is used, as discussed in Sec. 4.2.

As a study on the performances of subtraction methods, here the jet energy scale and groomed jet observables are measured in two different embedding scenarios and across three implementations of the constituent subtraction methods. Scales of the jet energies before and after embedding + subtraction are measured. To achieve so, jets that are clustered from background-subtracted events are matched to those at the generator level, such that their relative energy scales can be calculated. The results are shown in Fig. 4.9. The groomed substructure observables are compared between cases with and without thermal backgrounds (shown in Fig. 4.10). Note that an additional cut of $R_g > 0.1$ is applied.

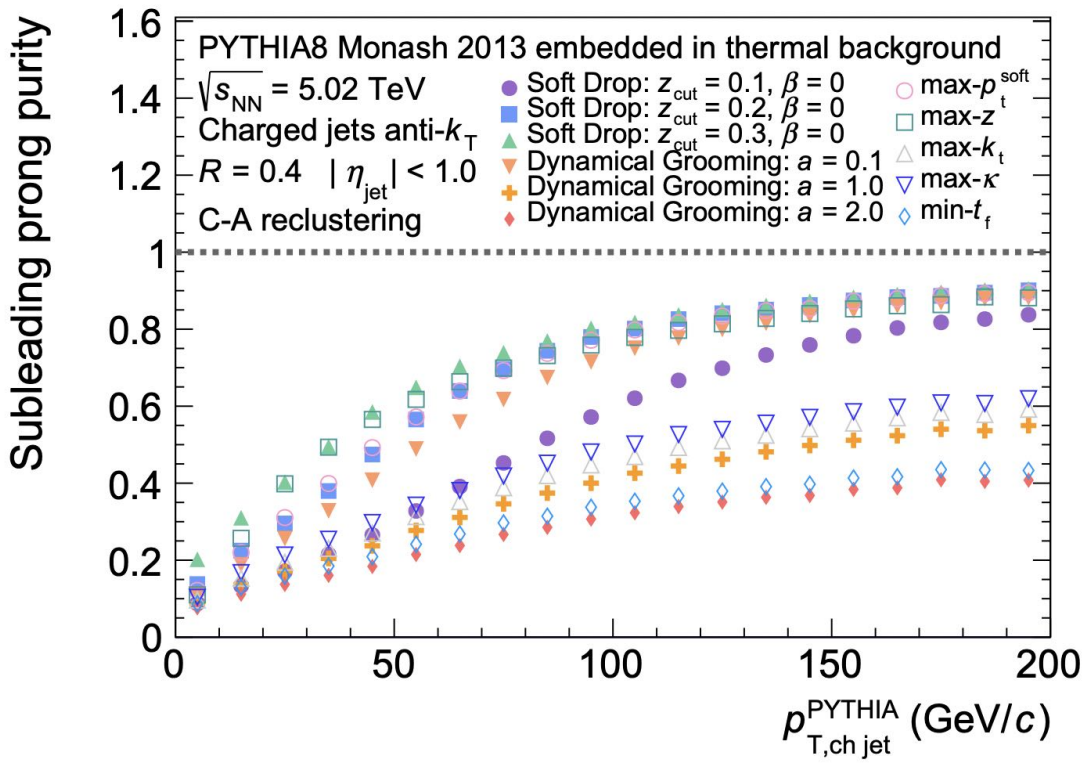

Figure 4.8: Subleading prong tagging purity as a function of jet p_T, measured for various groomers. The purity includes only the cases where the subleading prong is correctly tagged as the same one in the embedded scenarios.

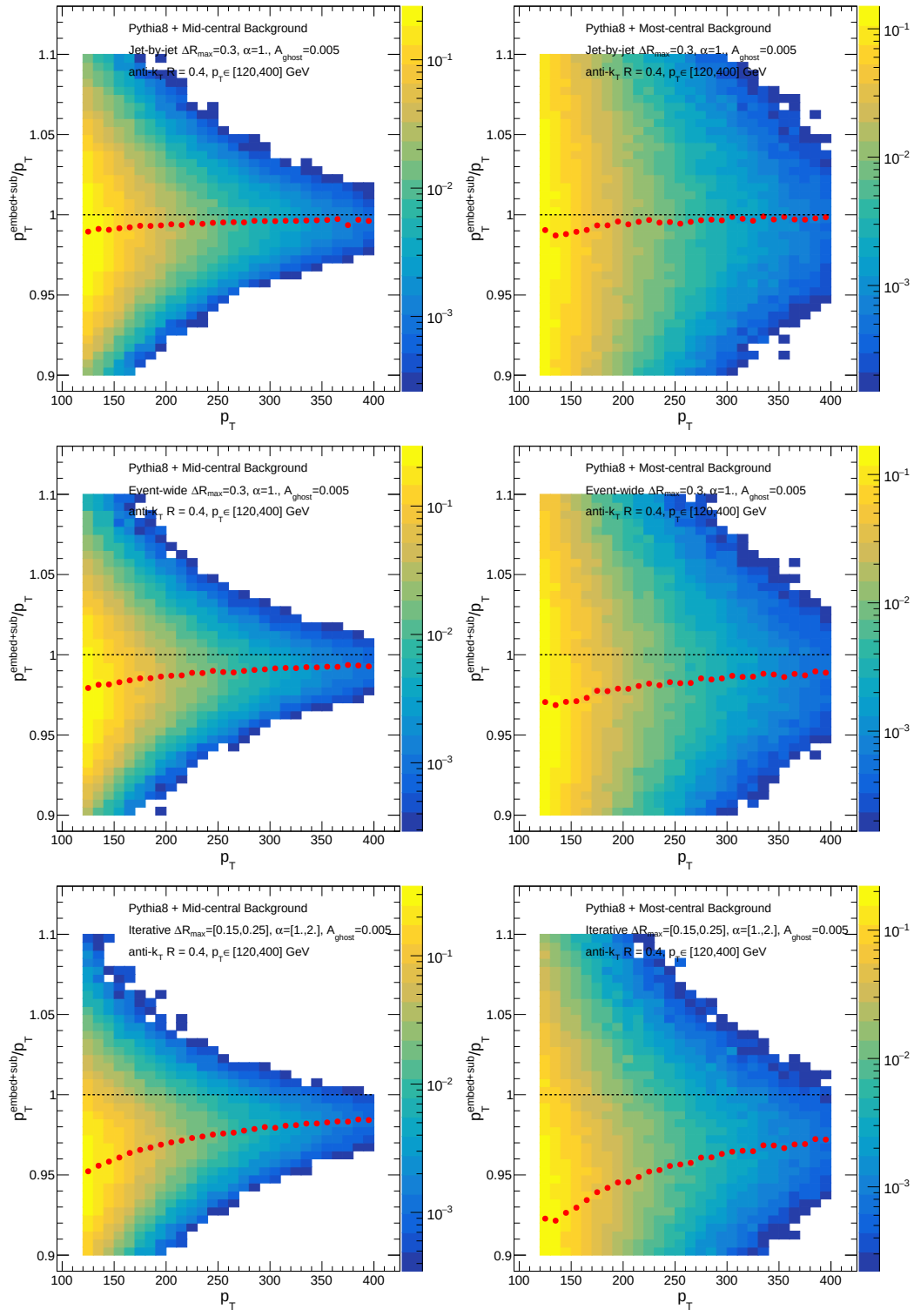

Figure 4.9: Jet energy scale, a 2D distribution of $p_T^{embed+sub}/p_T$ versus p_T where p_T is the jet momentum at generator level, $p_T^{embed+sub}$ is the momentum of the matched jet after embedding + subtraction. Red points are taken from a Gaussian fit for each p_T bin. Three different subtraction methods, the "jet-by-jet" (top), the "event-wide" (mid) and the "iterative" (bottom) are measured with mid-central (left) embedding and most-central (right) embedding.

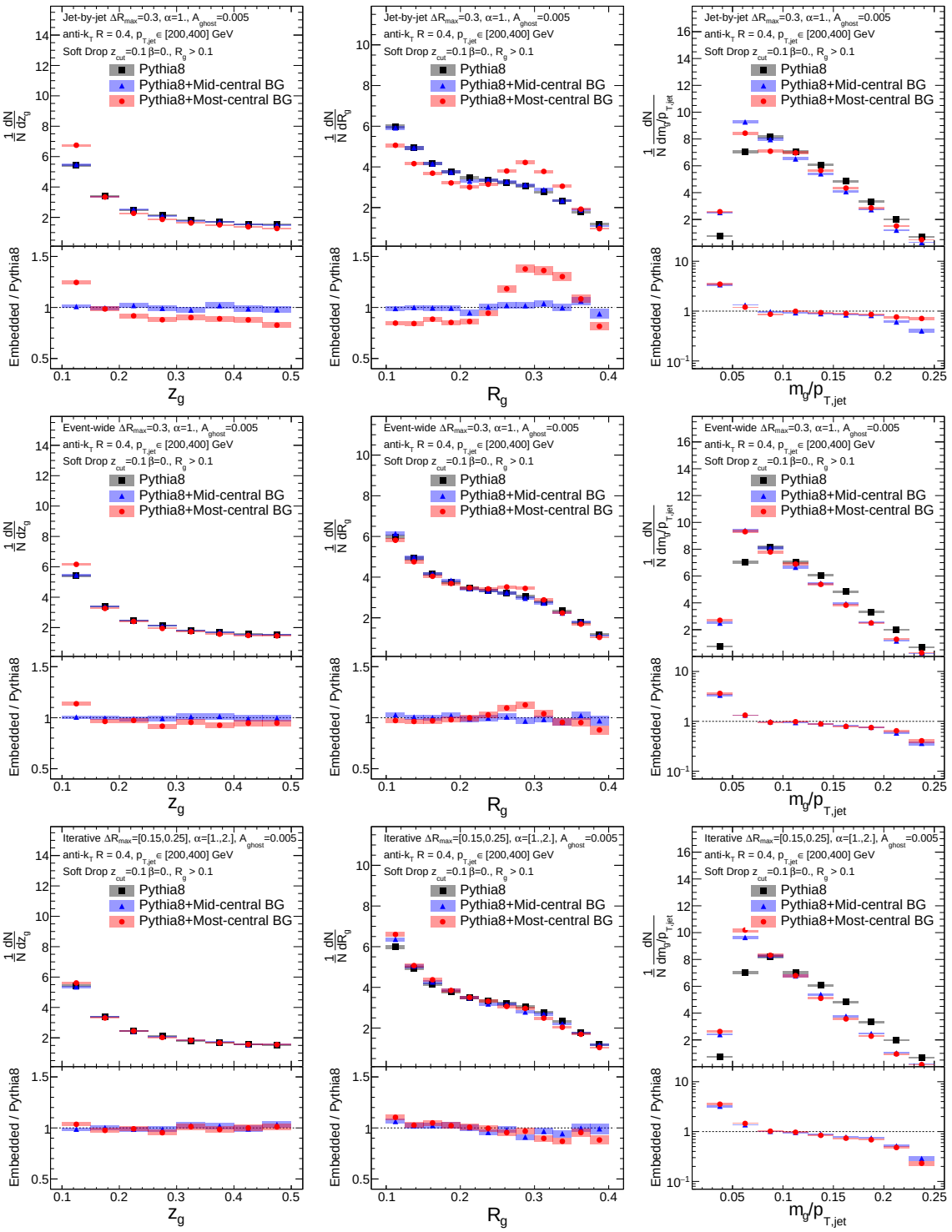

Figure 4.10: Groomed Substructure observables before/after embedding, with constituent subtraction applied. Comparison across three implementations, the "jet-by-jet" (top), the "event-wide" (mid) and the "iterative" (bottom) are shown.

The "jet-by-jet" method fails to subtract background contamination at larger distances relative to the jet axis, especially at high multiplicity embedding, leading to excesses at $R_g \approx 0.3$. On the contrary, both the "event-wide" and the "iterative" methods work well in constraining background effects on groomed substructure observables, while the "iterative" method tends to subtract more around the jet border, which leads to deletions at larger R_g. In addition, the "iterative" method is a little bit more aggressive than the "event-wide" method, showing an over-subtraction at smaller jet p_T. Based on the measurements on the jet energy scale and the groomed substructure observables, a choice of the subtraction method can be made. The "event-wide" constituent subtraction is preferred and is used in later studies.

4.3.2 Sensitivity to jet quenching

With simulations of the dijet events and the thermal backgrounds in heavy ion collisions, groomed substructure observables can be compared between quenched jets and vacuum jets, which will be shown in this section. PYTHIA events and JEWEL events are embedded into thermal backgrounds, after which a background subtraction method is applied to the embedded events. As discussed in previous sections, the "event-wide" constituent subtraction method is used, with free parameters $\Delta R_{max} = 0.3$, $\alpha = 1$ and $A_{ghost} = 0.005$. The anti-k_T algorithm is then applied to background-subtracted events with a definition of a jet radius $R = 0.4$. The constituents of each anti-k_T jet then undergo the C/A algorithm, resulting in an angular-ordered substructure, on which the soft drop declustering is performed with free parameters $z_{cut} = 0.1$, $\beta = 0$. The distributions of the groomed substructure observables are shown in Fig. 4.11, which can show a strong modification on the hard splittings due to quenching with enhancement on wider (larger R_g) and more unbalanced splittings (lower z_g).

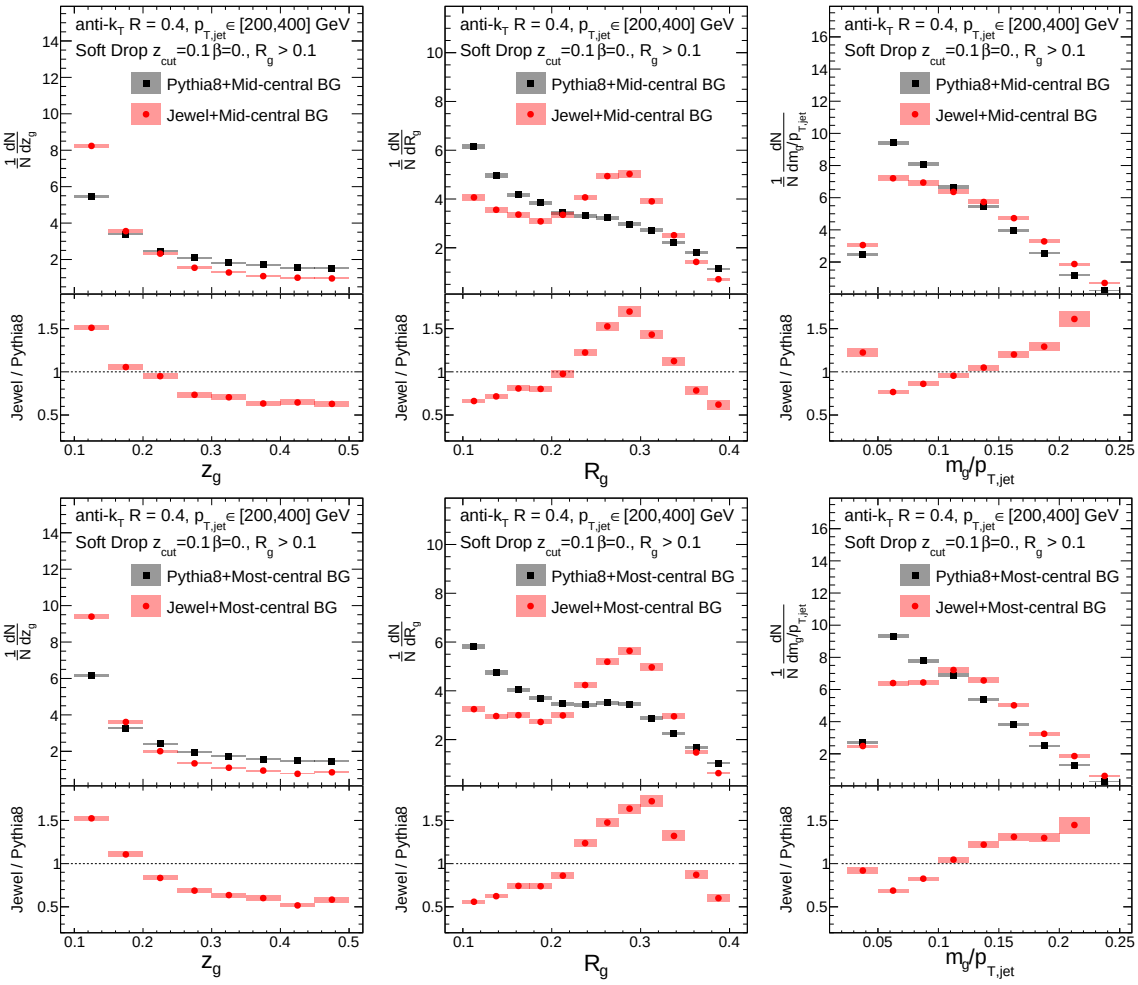

Figure 4.11: Groomed Substructure observables of the vacuum jets (PYTHIA) and the quenched jets (JEWEL), with constituent subtraction applied to embedded events. Dijet events are embedded into the mid central backgrounds (upper) and into the most central backgrounds (lower).

4.4 The Primary Lund Plane

The QCD splitting function in the soft and collinear litmit (Eq. 2.14) can be refactored in double-logarithmic format,

$$dP_{i\to ig} \simeq \frac{2\alpha_s C_i}{\pi} d\,ln z\theta\, d\,ln\frac{1}{\theta}. \tag{4.3}$$

Hence, the plane spanned by $lnz\theta$ and $-ln\theta$ is uniformly populated by such splittings, with a weight of $2\alpha_s C_i/\pi$. The primary emission off an emitter, described by kinematics variables z and θ, thus can be represented as a single point on the plane, namely the primary Lund plane [74]. As a semi-inclusive measurement of the QCD splitting function, the groomed substructure variables z_g and R_g are closely related to the hard splitting of an energetic parton. Thus, the primary Lund plane comes as a novel tool, providing a description of the phase space of the splitting.

In heavy ion collisions with the presence of QGP, modifications may happen on the QCD splitting function, leading to a modified phase space of the splitting. This can be tested by the measurement of the primary Lund plane density, defined as

$$\rho(z_g, R_g) = \frac{1}{N_{jet}} \frac{d^2N}{d\,ln(z_g R_g)\, d\,ln(1/R_g)}, \tag{4.4}$$

where z_g and R_g are taken from the soft drop declustering as described in previous sections.

A region of interest on the primary lund plane is accessible with the help of the soft drop groomers, as constraints are applied to splitting kinematics. Such kind of selection can be made through configurations on the free parameter set (z_{cut}, β) of the soft drop groomer. For illustration, the primary Lund plane with regions of interest selected by the soft drop groomers of three different configurations is shown in Fig. 4.12.

Here, the primary Lund plane density is measured with PYTHIA and JEWEL events embedded into thermal backgrounds, which respectively simulate vacuum jets and quenched jets in the presence of thermal

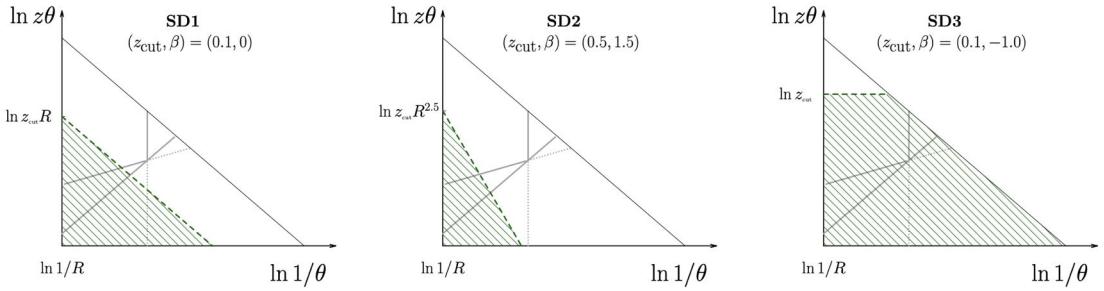

Figure 4.12: The primary Lund plane, on which regions of interest are selected by the soft drop groomers. Shaded areas correspond to kinematic regions that are groomed away. Figure from [57].

backgrounds, as shown in Fig. 4.13. It can be seen, vacuum jets simulated by the PYTHIA event generator populates the primary Lund plane uniformly, while quenched jets simulated by JEWEL show a modified phase space of splitting, with enhancements on the lower-left corner corresponding to wider and softer splittings.

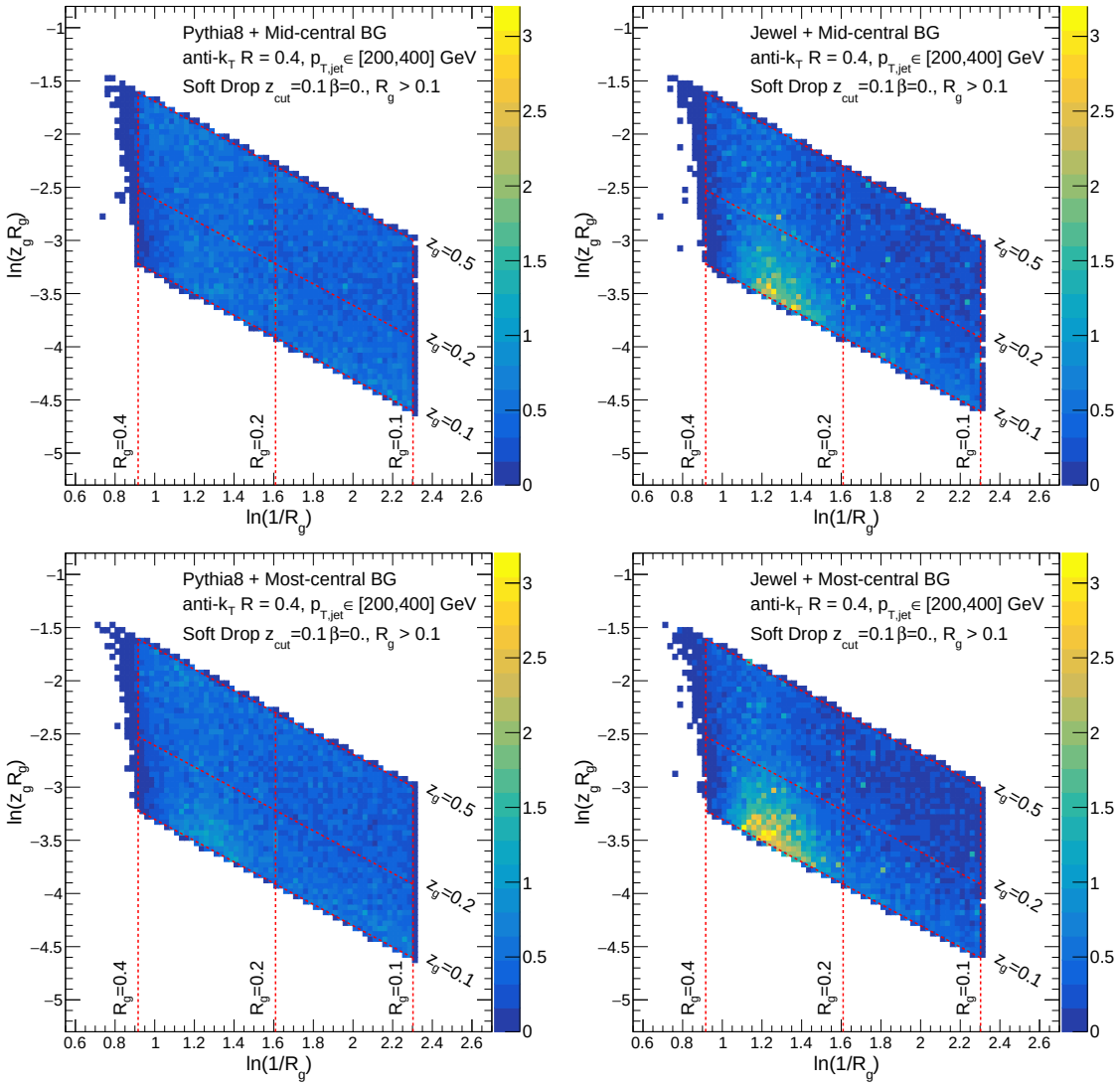

Figure 4.13: Plots pf the primary Lund plane density, filled with jets clustered from background-subtracted PYTHIA events (left) and JEWEL events (right), embedded into the mid-central backgrounds (upper) and the most-central backgrounds (lower). Soft drop is applied in order to access the substructure observables of jets. Dashed lines in red correspond to kinematic limits.

CHAPTER 5

Machine Learning

The previous chapters aim to introduce a basic understanding of the jet substructure from an experimental perspective but are still theory-driven. These Monte Carlo simulations become a start point for testing a new modern data analysis tool, that is, machine learning. Such data-driven techniques allow for new approaches to the analysis of jet substructure and uncovering new patterns in nature. Enthusiasts have been leading in the development of machine learning applications in particle physics[1].

As the main topic of this thesis, a machine learning approach to the identification of quenched jets is developed, which will be introduced in this chapter. A brief description of the convolutional neural network (CNN) will be given, after which the recurrent neural network (RNN) will be discussed, together with the description of a sequence-based feature engineering method designed specifically for jets. A supervised machine learning strategy is applied to the classification problem for the purpose of distinguishing quenched jets from vacuum jets. In addition, the robustness of trained models will be discussed.

5.1 Convolutional Neural Network

The convolutional neural network (CNN) (shown in Fig. 5.1) could be the first artificial neural network for a novice to begin with. It is commonly used to process visual images or videos. Its architecture involves three modular parts, often named layers, which are collections of neurons operating together. The three layers are typically the convolutional (Conv) layer, the pooling layer, and the fully connected (FC) layer.

[1] A living review of machine learning for particle physics: https://iml-wg.github.io/HEPML-LivingReview/
[1] Firgue from https://en.wikipedia.org/wiki/Convolutional_neural_network.

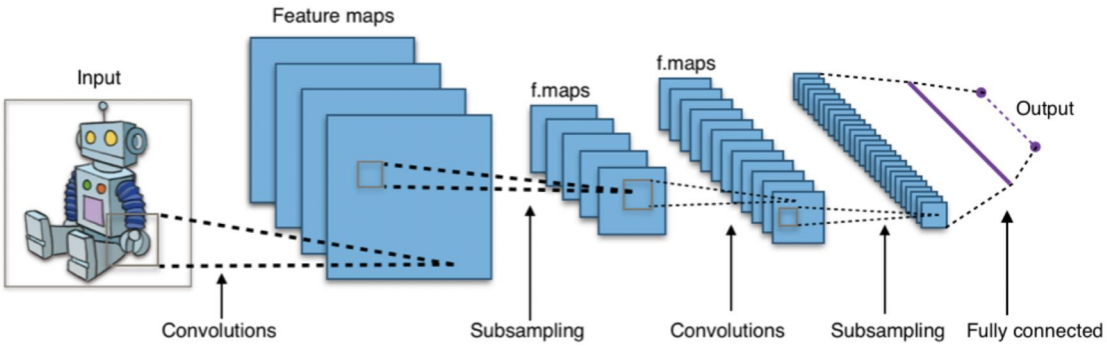

Figure 5.1: Typical CNN architecture.[1]

5.1.1 Architecture

The convolution (conv) layer is the core building block of a CNN, with learnable parameters in filters. A filter, also named as a kernel, is a matrix of weights (and biases). During the convolution operation, each filter is convolved across the area of the input matrix-shaped volume, computing the dot product between the filter parameters and the input elements, producing a 2-dimensional activation map of that filter. For illustration, the convolution operation is shown in Fig. 5.2.

In practice, a few parameters[3] control the output dimension of the convolution operation. Taken Fig. 5.2 for illustration, the input matrix has a dimension of (5,5,1) as its height, width and depth. The typical value of the depth for visual inputs, such as images and videos, is three, corresponding to RGB channels. Around the input matrix, padding of size 1 is added with zeros (zero-padded) or other values in padded cells. The grey overlap region on the input matrix represents the convolution filter (kernel) with a size of (3,3) as its height and width. The dot product of the filter and the input matrix is represented as the grey cell on the output matrix in green, after which the filter moves by 1 row (or 1 column) and a new dot product is calculated. The parameter so-called "stride" controls how much distance in cells the filter should move after each convolution operation. The convolution layer eventually produces an output matrix of dimension (5,5,1) with the given input matrix.

[2]Figure from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.
[3]Details can be found: https://pytorch.org/docs/stable/generated/torch.nn.Conv2d.html.
[1]Figure from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

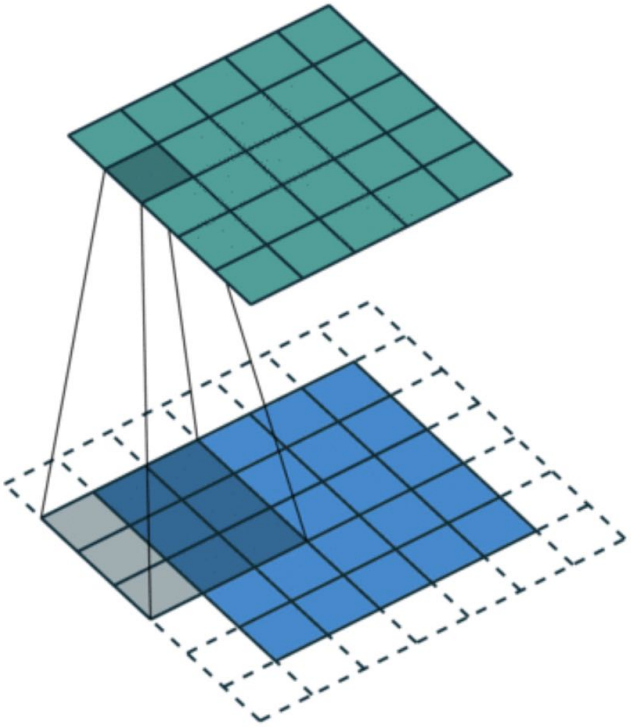

Figure 5.2: Illustration of the convolution operation in CNN. Cells in blue are the matrix-shaped inputs. Cells surrounded by dashed grey lines are paddings. The grey region on the input matrix represents where the convolution operation takes place. The green cells are the output matrix, on which the grey cell represents the dot product of the convolution operation.[2]

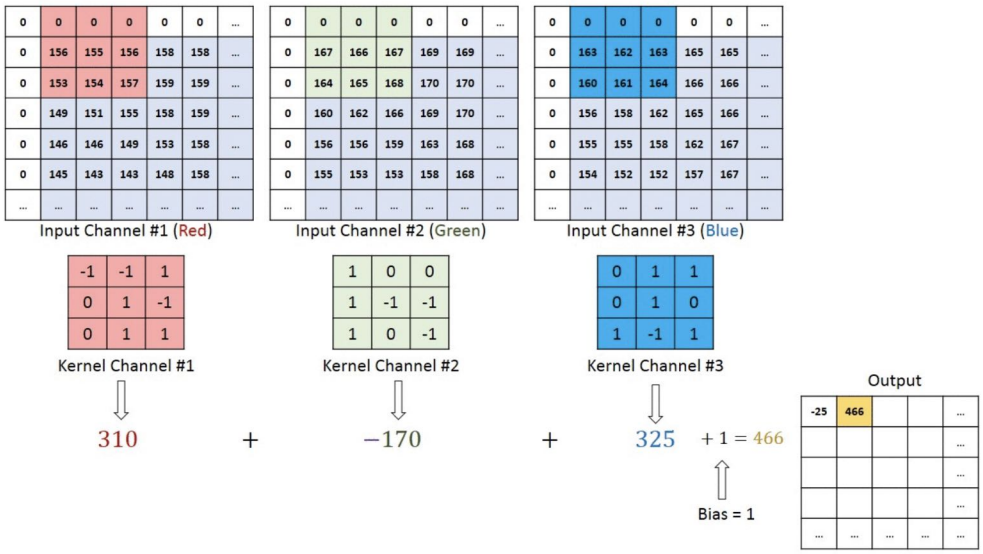

Figure 5.3: Illustration of the filters (kernels) of the convolution layer. Three filters work on three different input channels, corresponding to the RGB channels, and produce an output matrix.[1]

A typical CNN is often built with polling layers for the purpose of downsampling. The max-pooling is shown in Fig 5.4, which divides the input matrix into rectangles and outputs the max value of each sub-region. Such a downsampling method, as its name suggests, is able to significantly reduce the size of a matrix while keeping the most important features and the amount of calculation. Meanwhile, the pooling layers contribute to local translation invariance due to the fact that max-pooling operations are robust to small translations of the input matrix.

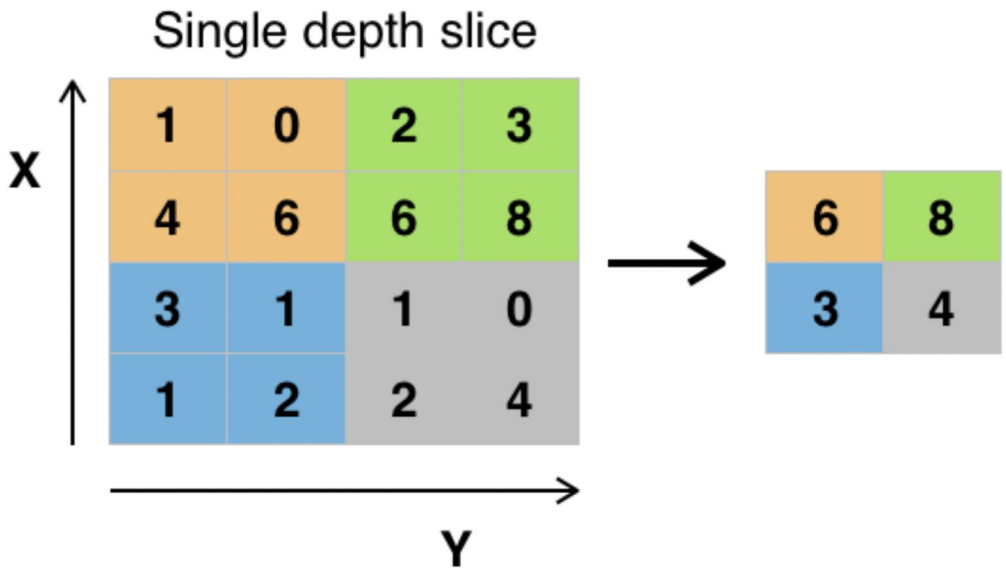

Figure 5.4: Illustration of the max-pooling layer.[1]

After several combinations of the convolution layers and the pooling layers, the final classification is done via fully connected (FC) layers (Fig. 5.5). In fully connected layers, the neuron applies a linear transformation to the input vector, done by typical matrix multiplication operations followed by bias offsets (either learnable or fixed). A non-linear transformation is then applied to the product through a non-linear activation function, for example, the softmax activation function. As shown in Fig. 5.5, all possible layer to layer connections are present. Thus the output values depend on every input value, with different weights and biases.

[1]Firgue from https://en.wikipedia.org/wiki/Convolutional_neural_network.
[1]Figure from https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b.

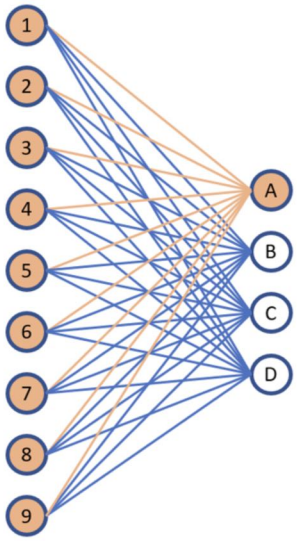

Figure 5.5: Illustration of the fully connected layer.[1]

5.1.2 Jet Images and preprocessing

In order to utilize the convolutional neural network (CNN) in jet physics, jet images are introduced. A jet image can be easily constructed by taking the constituents of a jet and discretizing its transverse energy into pixels on the $\eta - \phi$ plane (Fig. 5.6).

Preprocessing is commonly applied to images that are about to be processed by the convolutional network and are crucial, especially in jet physics. It helps the neural network to learn discriminating features and not learn the symmetries of space-time. Meanwhile, such a procedure can greatly improve performance and the required size of data used in training a neural network.

The most important step of the preprocessing, when it comes to jet physics, is alignment, involving several operations such as translation, rotation and reflection. Such operations are typically designed according to the characteristics of jets. In the translation operation, the center of the image is defined on the $\eta - \phi$ plane, implying that the center of a jet image can be different from the jet axis. In the rotation operation, jet images are typically rotated such that the primary features align along a specific axis. In the study of the W-jet

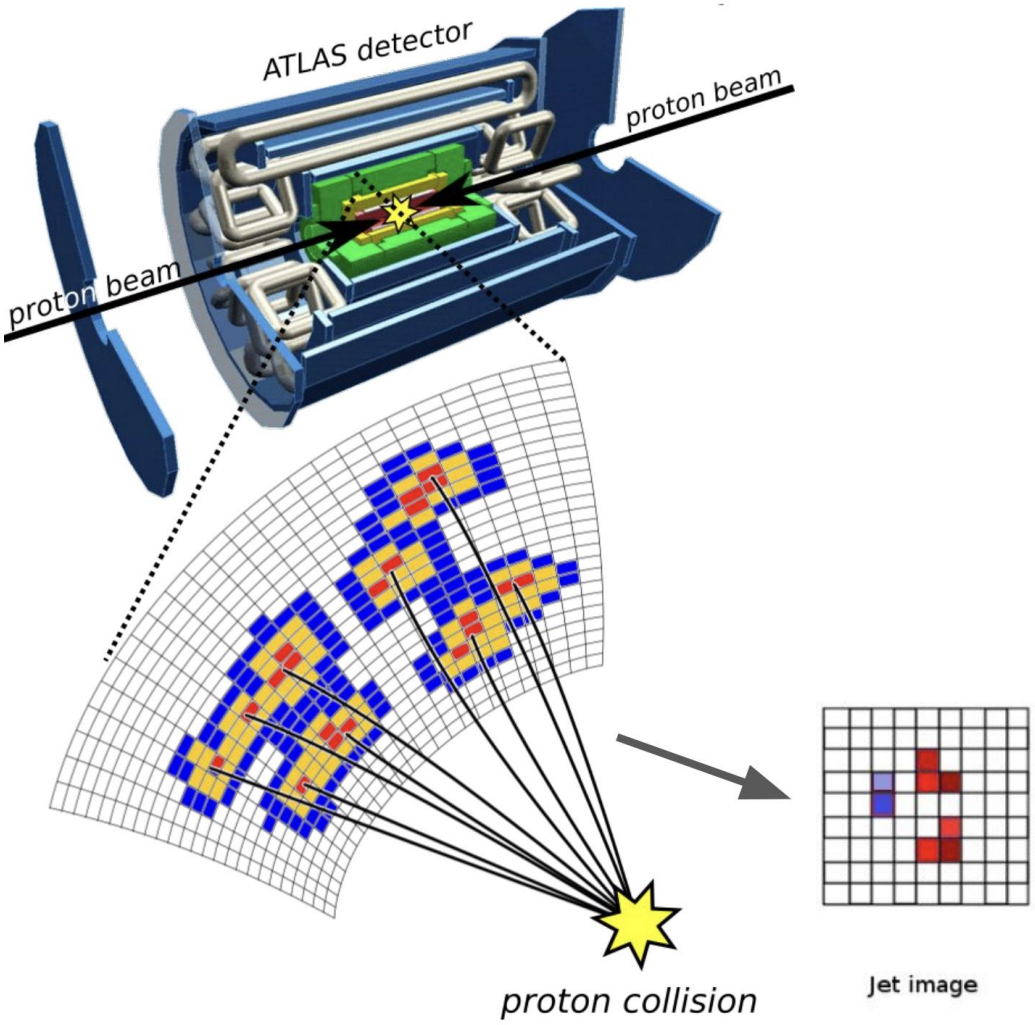

Figure 5.6: Construction of a jet image. Figure from The Third Machine Learning summer school (https://indico.cern.ch/event/613571/contributions/2617924/attachments/1496948/2329564/deepjets.pdf).

tagging [75, 76], such an axis is defined as the one starting from the center of the leading subjet to the center of the subleading subjet, based on the fact that a boosted W boson produces a jet with a "2-prong" structure. In contrast, the reflection operation can control on which side of the image the hardest radiation should appear. Fig. 5.7 shows the average jet-image for W jets before and after preprocessing, where there is a clear "2-prong" structure in the average jet image after preprocessing.

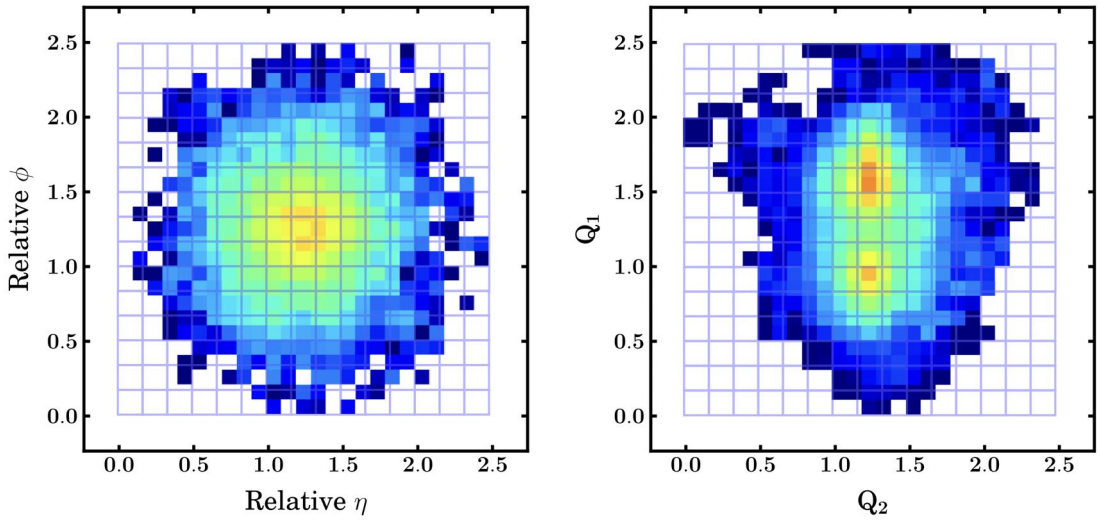

Figure 5.7: Average jet image for W jets, before (left) and after (right) preprocessing. Figures from Ref. [76].

5.1.3 Discussion

It can be seen from the convolution operation a filter scans over the whole image and subtracts the high-level information about whether a specific feature is present or not. Thus a convolution filter after training carries spatial information about the feature with discriminating power. This enables interpretation, empirically rather than analytically, from a well-trained CNN by the visualization of its convolution filters. For example, Fig. 5.8 shows the visualization of the convolution filters extracted from the first convolutional layer of a CNN, which is trained for the purpose of quark/gluon jets classification. Note that in the study (Ref. [77]), due to the lack of rotation operations in jet image preprocessing, many of the filters are simply rotations of other filters. This also implies that the rotation operation in the jet image processing helps to

reduce the number of filters, and eventually reduces computation complexity.

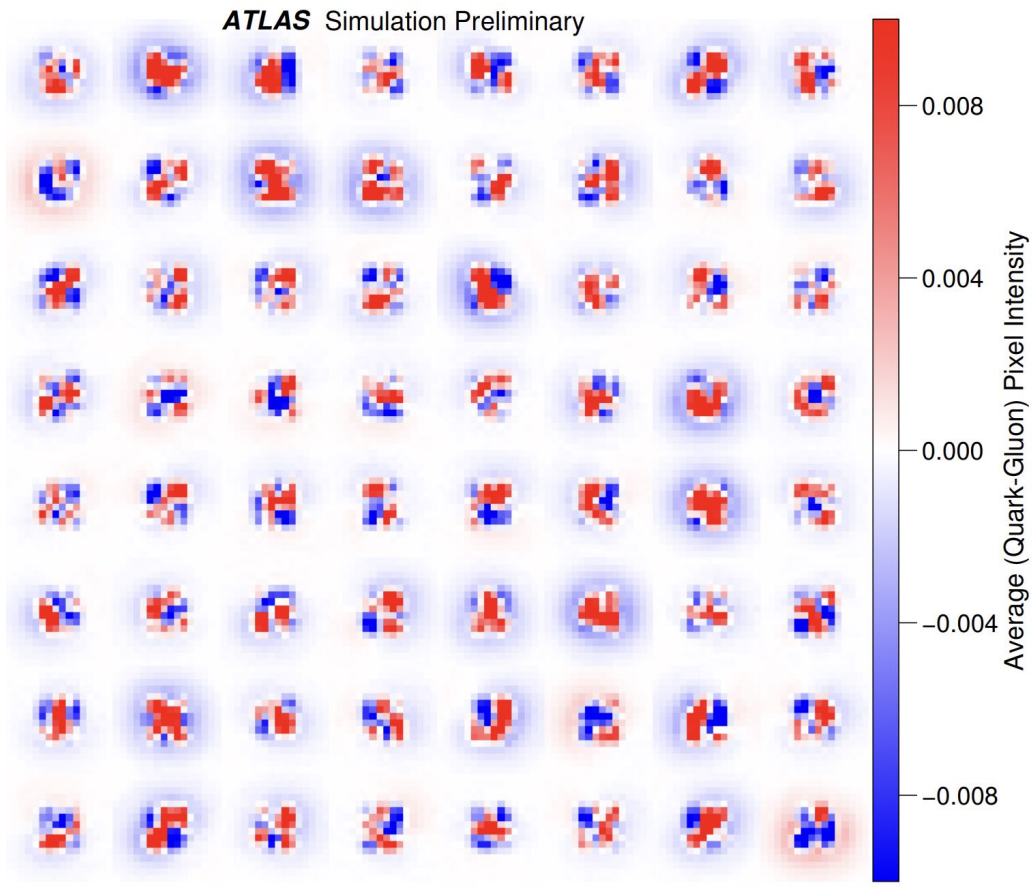

Figure 5.8: Average convolved filter differences for jet images, extracted from CNN trained for the purpose of quark/gluon jets classification. Figure from Ref. [77].

Even though the visualization of the convolution filters can show what features are important in the classification problem, analytical analysis of the filters is not practical (This claim is made based on the author's limited knowledge of machine learning). What is even worse, CNN has its limitation in learning freely rotated features. In other words, a feature after rotation simply becomes a new one. Thus a question arises, that is, whether an image after rotation can be categorized into the same class as it is. A simple experiment on the hand digit recognition problem is shown in Fig. 5.9, where a digit of "7" after rotation is wrongly classified. This simple experiment has shown that image preprocessing is necessary in achieving rotational invariance on classification.

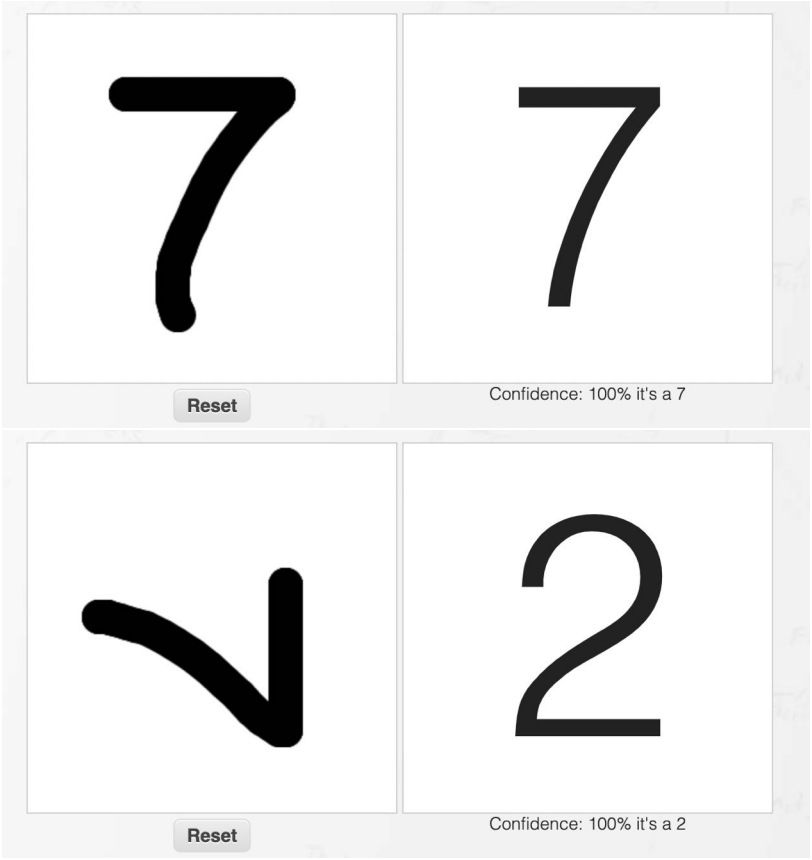

Figure 5.9: Example of the hand digit recognition with CNN. A digit of "7" is correctly recognized (upper), while a rotated "7" is wrongly classified as "2" (lower). Results from handwritten digit recognition playground (online): http://www.denseinl2.com/webcnn/digitdemo.html.

5.2 Recurrent Neural Network

In the later part of the thesis, the recurrent neural network (RNN) will be introduced[1] and is used. What makes it unique is that RNNs can use their internal state (memory) to process various length sequences of inputs. Thus it's capable of handling tasks such as natural language processing. An RNN, as its name suggests, has recurrent loops in the network, allowing information to persist and be discarded. An RNN can be viewed as multiple copies of the same network, each passing information to its successor, as shown in Fig. 5.10.

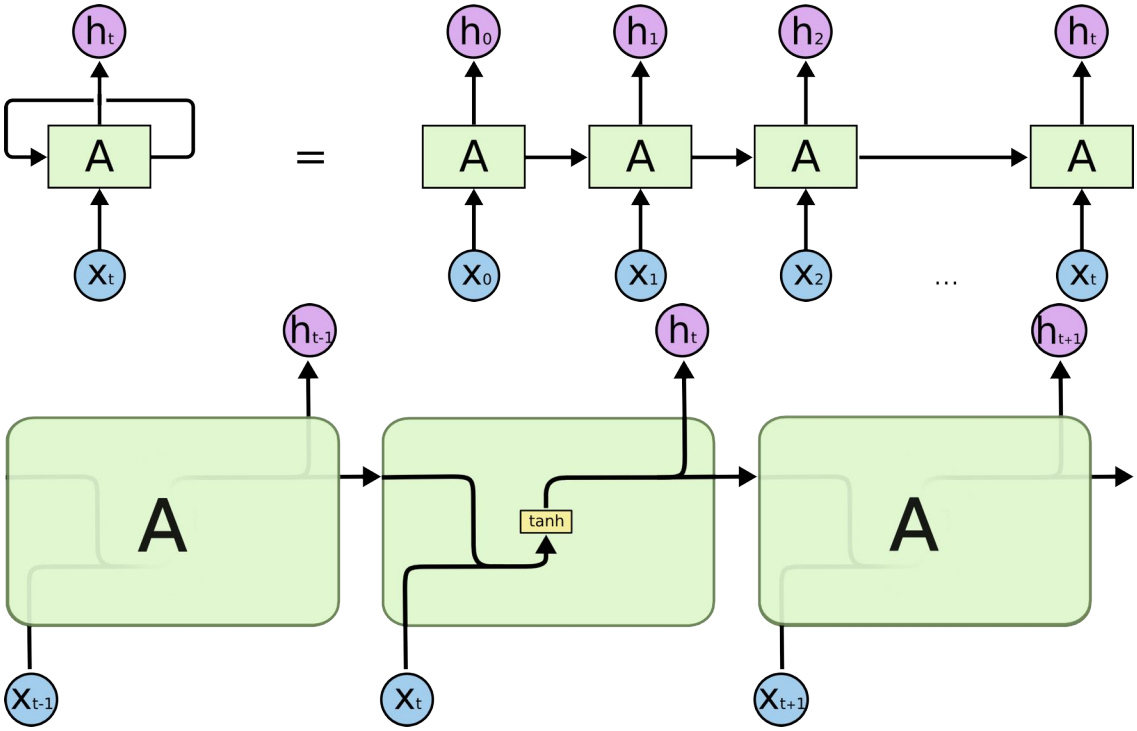

Figure 5.10: Illustration of the recurrent neural network (upper) with its repeating modules (lower). At step t, RNN takes input values x_t and produces output values h_t.

In theory, RNN should be able to deal with the long-term dependencies, for example, making predictions about the next value based on previous values. However, in practice, the vanishing gradient problem can happen in training. As a result, the long short-term memory (LSTM) network is invented, which has the advantage of relative insensitivity to gap length.

[1] Figures in this section are from https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

5.2.1 The long short-term memory (LSTM) Netwotk

LSTMs are explicitly designed to avoid the gradient problem, while maintaining the capability of learning long-term dependencies. The repeating module of an LSTM network has an improved design and is different from that of a traditional RNN, as shown in Fig. 5.11.

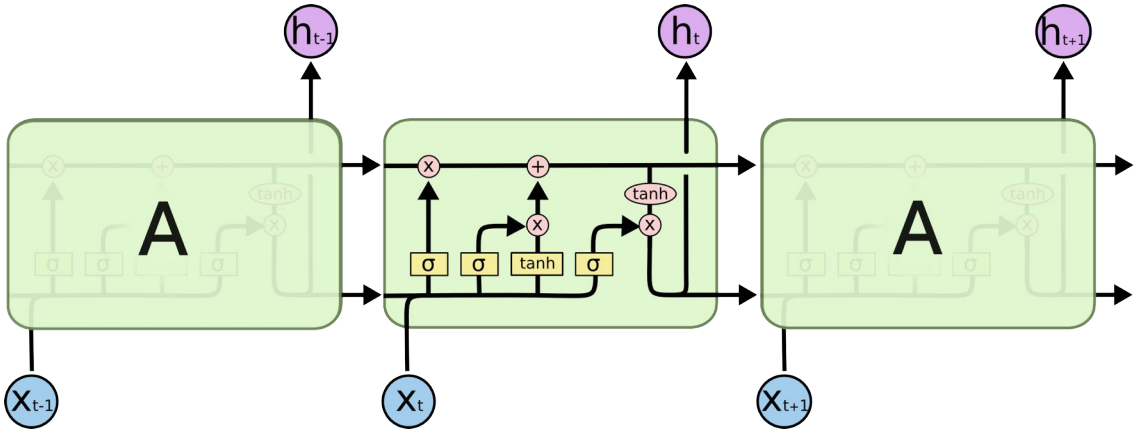

Figure 5.11: Repeating modules in a LSTM network.

The core idea behind the design of the LSTM network is the cell state, where information can be removed or added. In contrast, other structures (so-called "gates") are designed to regulate information flow in or out of the cell state. Details about the dynamics of the LSTM modules are beyond the purpose of this thesis but can be easily found in studies on the behaviors of RNN and the LSTM (Ref. [78, 79]). In the PYTORCH implementation, several LSTM layers can be used together to form a stacked LSTM, with each taking in outputs from the previous LSTM layer, computing outputs and feeding results to the next. Outputs from the LSTM network ($[h_0, \ldots, h_t, \ldots]$) contain output values h_t from the last layer of the LSTM for each t, with "hidden size" referring to the size of h_t. Details about other configurable parameters can be found from PYTORCH online manuals[1].

[1]Description about the LSTM implementation from Pytorch online manual: https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html

5.2.2 Jet Sequential Substructure Observables

In order to make use of the LSTM network, a feature engineering method is designed specifically for jet physics by making use of the jet substructure observables. As described in Section 4.2 and Section 4.3, the C/A declustering procedure reveals the angular-ordered pairwise substructure of a jet, on which a hard splitting can be located by the soft drop groomer.

Once the hard splitting is located, the declustering continues on the harder subjet recursively until a subjet can no longer be divided. This procedure is similar to the recursive soft drop [80], but without grooming on the subleading subjet of each splitting. Note that whether to remove the subleading subjet is not studied in this thesis. But the benefit of not doing so is that a reasonable number of the total splittings can be achieved from the declustering procedure.

Along the desclustering, a sequence of splittings can be found, with all the harder subjets forming the feature branch as labeled in Fig. 5.12. On the feature branch, each point represents a splitting with substructure

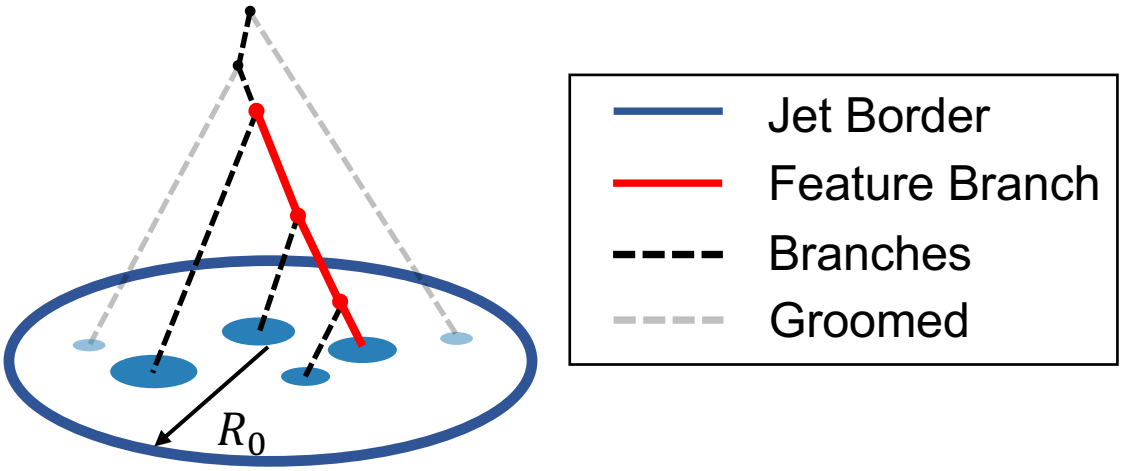

Figure 5.12: Feature branch selected from the declustering sequence of a jet.

observables describing its kinematics. The observables include, but are not limited to, the momentum fraction z, angular distance ΔR, perpendicular momentum $k_\perp$ and invariant jet mass m_{inv}, which are all calculated

from the pair of de-clustered subjets with

$$
\begin{aligned}
z &= \frac{min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}}, \\
\Delta R &= \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2}, \\
k_\perp &= min(p_{T,i}, p_{T,j}) * \Delta R, \\
m_{inv} &= \sqrt{(E_i + E_j)^2 - (\mathbf{p}_i + \mathbf{p}_j)^2}, \\
x_t &= [z, \Delta R, k_\perp, m_{inv}],
\end{aligned}
\tag{5.1}
$$

where i, j denote the two subjets at the de-clustering step t. In this way, sequential vectors $[x_0, \ldots, x_t, \ldots]$ are extracted along the feature branch, and are then used in the training of a neural network.

A similar feature engineering method is utilized in this study (Ref. [81]), where Lund coordinates ($lnz\theta$, $-ln\theta$) are extracted from the C/A declustering sequence of jets and are used in the training of an RNN. Ideas are the same, as sequential observables obtained from jet declustering can describe the kinematics of parton showering. Note that the Lund coordinates form a subset of the feature vector x_t.

5.3 Supervised Machine Learning

In order to apply the machine learning method in jet physics, the best practice is to start from a simple classification problem. In this thesis, the machine learning approach to the identification of quenched jets against vacuum jets is studied only with simulations. A similar study has been done (Ref. [81]), focusing on different neural network architectures and their performances, while thermal backgrounds or event biases are not taken into consideration.

As described in Chapter 3, jets are clustered from background-subtracted events that are mixed from dijet events and thermal background events. Jets associated with the PYTHIA event generator are labeled with 0, referring to as the vacuum (non-quenched) jets or the negative class. In contrast, jets associated with the

JEWEL event generator are labeled with 1 and are referred to as the quenched jets or the positive class. In addition to a label indicating whether the jet is quenched, each jet has a weight that is scaled from the weight of the dijet event given by its generator. Such event weight is typically proportional to its cross-section. In this thesis, the weight of a jet is calculated as the event cross-section divided by the total, meaning that the weights of the whole class sum up to 1. Note that by doing this, the relative weights instead of the absolute cross-sections are used in the next step, that is, the loss calculation. Weights are used for compensation due to biased selection towards high p_t dijet events at the generator level. It should be pointed out that simulations with biases and weights are common practices in Monte Carlo simulations as more samples from interested phase space can be generated to gain statistical benefits. In this thesis, a machine learning approach suitable to such simulations is designed, with weights taking roles in the loss calculation, which will be discussed in the next section. As for real experimental data, weights can be simply downgraded to 1.

A total number of 400k embedded events are simulated with PYTHIA and are divided into halves. Inclusive jets over 120 GeV clustered from events of each half are organized into a training set and a validation set. This procedure is also applied in simulations with JEWEL. A jet p_T cut of $p_T > 200$ GeV and a cut on the jet substructure observable $R_g > 0.1$ are applied. In addition, jets with only one constituent are discarded. The total numbers of jets with and without cuts are listed in Tab. 5.1.

No. of Jets	Training Set	Validation Set
PYTHIA+Mid-central	43203 / 314814	43034 / 314533
PYTHIA+Most-central	42535 / 310332	42272 / 310276
JEWEL+Mid-central	52908 / 302280	52500 / 302707
JEWEL+Most-central	52954 / 298675	52967 / 298876

Table 5.1: Description of datasets used in training and validation, w/ and w/o cuts.

5.3.1 Training of a neural network

A neural network has built-in learnable parameters (weights and biases) and is able to map high dimensional input data to a vector of values. In this way, predictions on the categories of the inputs are made. But a

neural network is not able to achieve this on initialization. In the training process, those learnable parameters are adjusted on the guidance of a loss function.

The loss function should work in a way that can quantify how well a neural network works in making correct predictions. The commonly used loss functions in binary classification are the mean square error (MSE) loss function[1] and the binary cross entropy (BCE) loss function[2], described by

$$
\begin{aligned}
l_{MSE} &= \frac{\sum\limits_{batch} \omega_i * (x_i - y_i)^2}{\sum\limits_{batch} \omega_i}, \\
l_{BCE} &= \frac{\sum\limits_{batch} \omega_i * (y_i log x_i + (1 - y_i) log(1 - x_i))}{\sum\limits_{batch} \omega_i},
\end{aligned}
\tag{5.2}
$$

where x_i, y_i, ω_i refer to the predictive label, the truth label and the input weight of the sample with index i. As a typical practice, the loss is summed over samples in a batch with its size determined by the parameter "batch size".

Batched samples taken by RNN are required to have the same lengths. In this thesis, post-sequence paddings are added to the samples in a batch, such that samples have equal lengths. Discussion about padding is in Appendix A. The mean square error (MSE) loss function is used. Discussion about the two loss functions is in Appendix B.

The training of a neural network involves the following steps:

1. Randomly select a number of samples from both the positive and the negative classes to form a batched input.

2. The neural network takes the batched input and gives predictive label for each sample in the batch.

3. Calculate the training loss (MSE or BCE) based on the predictive label from the neural network and

[1] Details about MSE can be found: https://pytorch.org/docs/stable/generated/torch.nn.MSELoss.html.
[2] Details about BCE can be found: https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html.

the truth label.

4. Adjust the learnable parameters of the neural network, such that the training loss is decreasing over training steps. For illustration, the training loss as a function of training steps is plotted in Fig. 5.13.

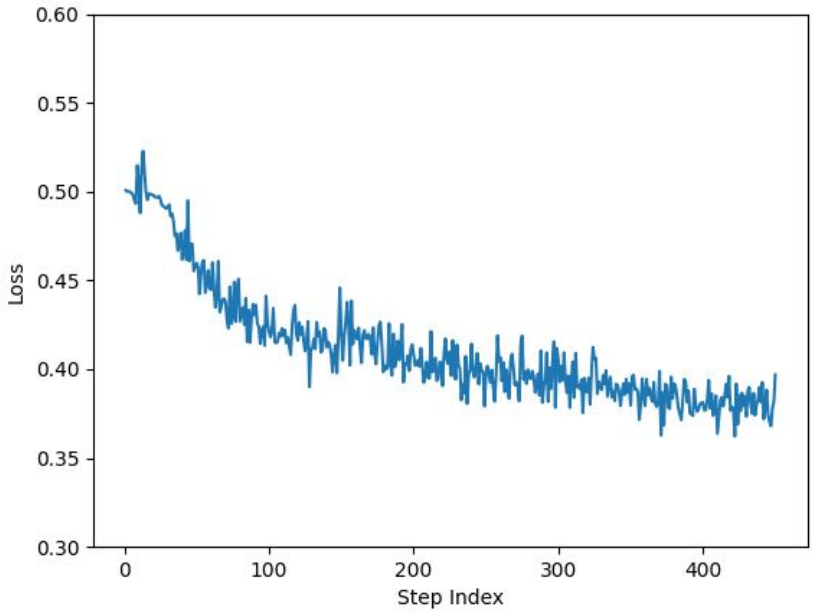

Figure 5.13: Example of training loss (MSE) decreasing in the training of a neural network.

5.3.2 Validation and Hyper-tuning

In machine learning, validation is referred to as the procedure where a trained model is assessed with a validation dataset. The validation loss (MSE or BCE) is calculated with the given validation dataset as the measurement of the performance of a trained neural network. The validation dataset differs from the training dataset to guarantee the best model picked in the validation has no way to memorize the training dataset.

Validation loss is used in the so-called hyper-tuning of a neural network. Unlike the learnable parameters, which can be adjusted in the training process, hyper-parameters can not be adjusted during the training. Thus, hyper-tunning refers to the strategy of optimizing the neural network by selecting a hyper-parameter set. In

this thesis, hyperparameters related to the architecture of a neural network and its training are tuned with the HYPEROPT [82] package.

The LSTM neural network is built with stacked LSTM layers and full-connect layers. The hidden size of the LSTM layer, as described in Section 5.2.1, is the size of the output vector h_t at step t. The output of the last step from the top LSTM layer is directed to two full-connect layers. Thus the input dimension of the first full-connect layer is equal to the hidden size of the LSTM neural network. The output dimension of the first full-connect layer is equal to the input dimension of the second. Both the input and output dimensions of the first full-connect layer, with the former being equal to the hidden size of the LSTM layer, are the hyper-parameters defining the architecture of the neural network.

During the training, the training dataset is used multiple times, with each named as an epoch. In an epoch, the dataset is shuffled in order to increase the randomness. The training process stops once the total training epochs reach a maximum. The total training steps in one epoch can then be calculated as the size of the training dataset divided by the batch size. After each step, the neural network will adjust its learnable parameters, where the learning rate controls how much the parameters should be changed. After each epoch, the learning rate reduces based on a given decay factor. The number of maximum training epochs, the batch size, the learning rate and its decay factor are the hyper-parameters related to the training process.

We scan over the hyper-parameter search space 50 times and arrange 3 times of training for each hyper-parameter set. The trained model with minimal validation loss is selected as the best. The distribution of validation loss for different hyper-parameter options is displayed using violin plots, showing both the probability density of validation loss and its quartiles, as shown in Fig. 5.14. The full hyper-parameter sets and their optimal values are listed in Tab. 5.2.

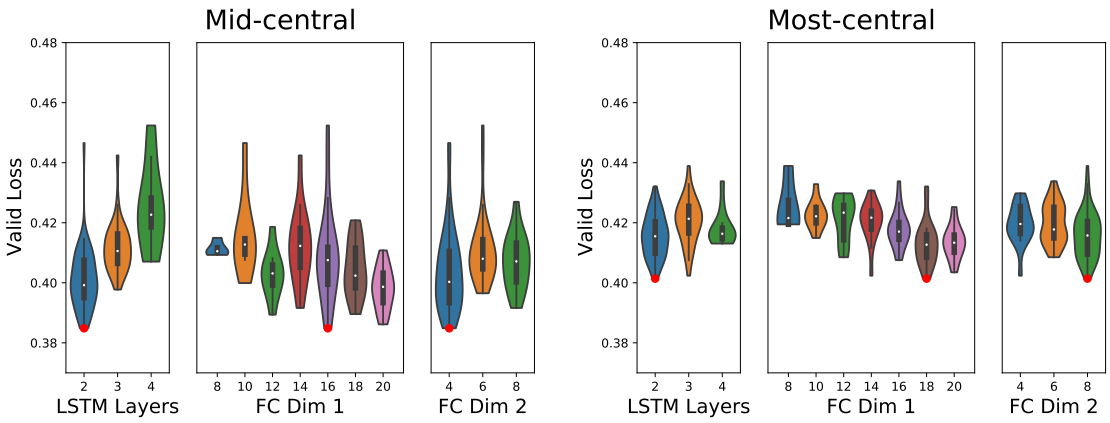

Figure 5.14: Distribution of validation loss for different architecture configurations, for the mid-central mixing (left) and the most-central mixing (right) respectively. The best models are indicated in red. The outer possibility density contour cuts off at min/max values. The inner white point represents the median value. The inner thick bar shows the first and the third quartiles.

Parameters	Mid-central	Most-central
No. of LSTM Layers	2	2
FC Dim 1	16	18
FC Dim 2	4	8
No. of Epochs	50	45
Batch Size	10000	18000
Learning Rate	0.0250	0.0371
Decay Factor	0.9689	0.9896

Table 5.2: List of hyper-parameters related to neural network architecture and the training process and their optimal values after hyper-tuning.

5.4 Robustness

A trained neural network can be used as a classifier, making predictions about how quenched a jet is on a jet-by-jet basis. It is noticed the distributions of raw LSTM output from trained neural networks exhibit a non-deterministic behavior in the metric. During the training, samples of the negative class are labeled with 0, while samples of the positive class are labeled with 1, without intermediate labels in between. Thus the metric that quantifies the quenching effect suffers from the stochastic nature of machine learning, rooted in the randomness of selecting samples to form a batch. A calibration method is designed to solve the non-determinism issue.

In binary classification, the predictive labels are either positive (P) or negative (N). Thus four possible situations could happen as a combination of the predictive labels and the true labels, as shown in Fig. 5.15. If the prediction is P and the actual value is also P, then it is called a true positive (TP); however, if the actual value is N then it is said to be a false positive (FP). Conversely, a true negative (TN) occurs when both the prediction outcome and the actual value are N, and false negative (FN) is when the prediction outcome is N while the actual value is P. With a given threshold, The true positive rate (TPR) and the false positive rate (FPR) hence are defined as,

$$TPR = \frac{TP}{TP+FN},$$
$$FPR = \frac{FP}{FP+TN}. \tag{5.3}$$

Considering that the selection of thresholds for classification affects the predictive labels, a curve of TPR

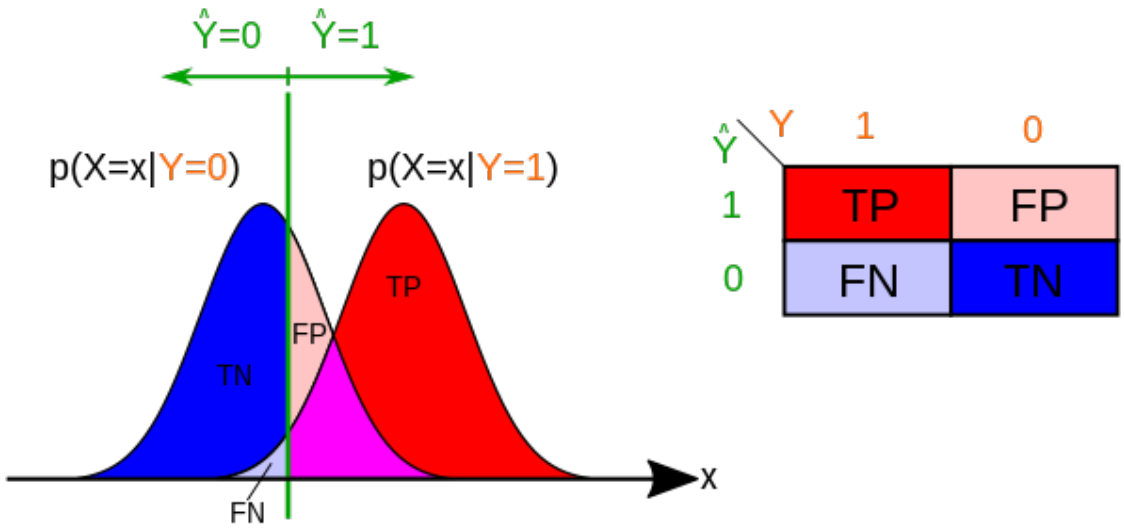

Figure 5.15: Outcomes in binary classification, where x is the "score" given by the classifier, Y is the true label and $\hat{Y}$ is the predictive label inferred from x and a threshold.

versus FPR at various thresholds can be plotted, as shown in Fig. 5.16. Such plot is named as the receiver operating curve (ROC), with the area under the curve (AUC) showing the performance of the classifier.

A universal threshold for the binary classification across all trained neural networks is not achievable. In order to solve the non-determinism issue, a calibration method is designed with the help of the ROC curve. To better explain how thresholds can be determined from the ROC curve, three points are plotted, with each

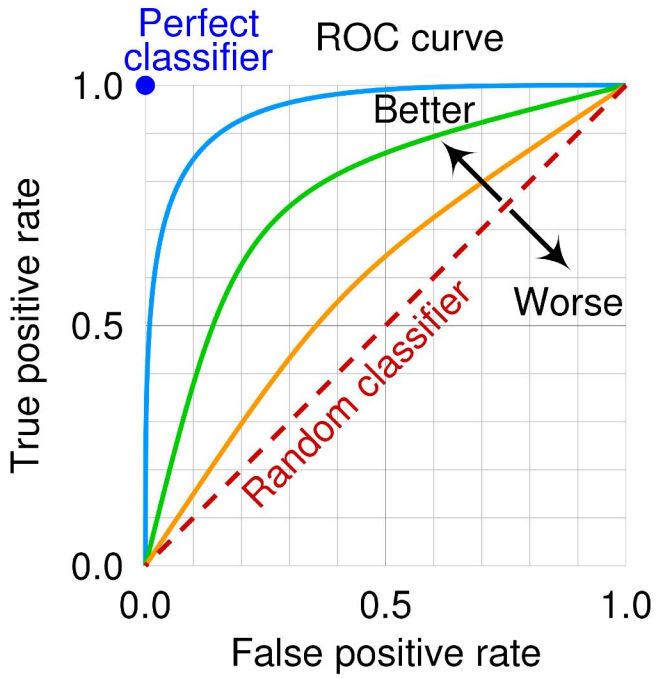

Figure 5.16: The receiver operating curve (ROC) in binary classification, with a bigger area under curve (AUC) being a better classifier.

associated with a specific TPR or FPR, as shown in Fig. 5.17. The threshold associated with $TPR = 0.4$ (marked in red) can divide samples of the positive class into two subsets at a $60/40$ ratio. In contrast, the threshold associated with $FPR = 0.6$ (marked in blue) can divide the samples of the negative class into two subsets at a $60/40$ ratio. The point marked in green is the closest point to the upper left corner of the ROC curve, with its inferred threshold being commonly selected for binary classification.

The calibration method is designed such that samples of the negative class are uniformly distributed along with the newly established metric. Meanwhile, the relative quantitive relationships between samples' raw LSTM output are kept in order. The new metric is bound to FPR; meanwhile the threshold on the raw LSTM output corresponding to a specific FPR value can be determined from the ROC curve. The distribution of the raw LSTM output from trained neural networks and their values after calibration are shown in Fig. 5.17. It can be seen that such a calibration method produces almost identical distributions of the calibrated LSTM values for three well-trained classifiers.

Choice of the threshold is based on the purpose of division. In later analysis, the threshold associated with $TPR = 0.4$ is used to divide the positive class into two subsets, one subset of which all the samples pass the threshold and another. The differences between samples of the two subsets are shown in the next chapter.

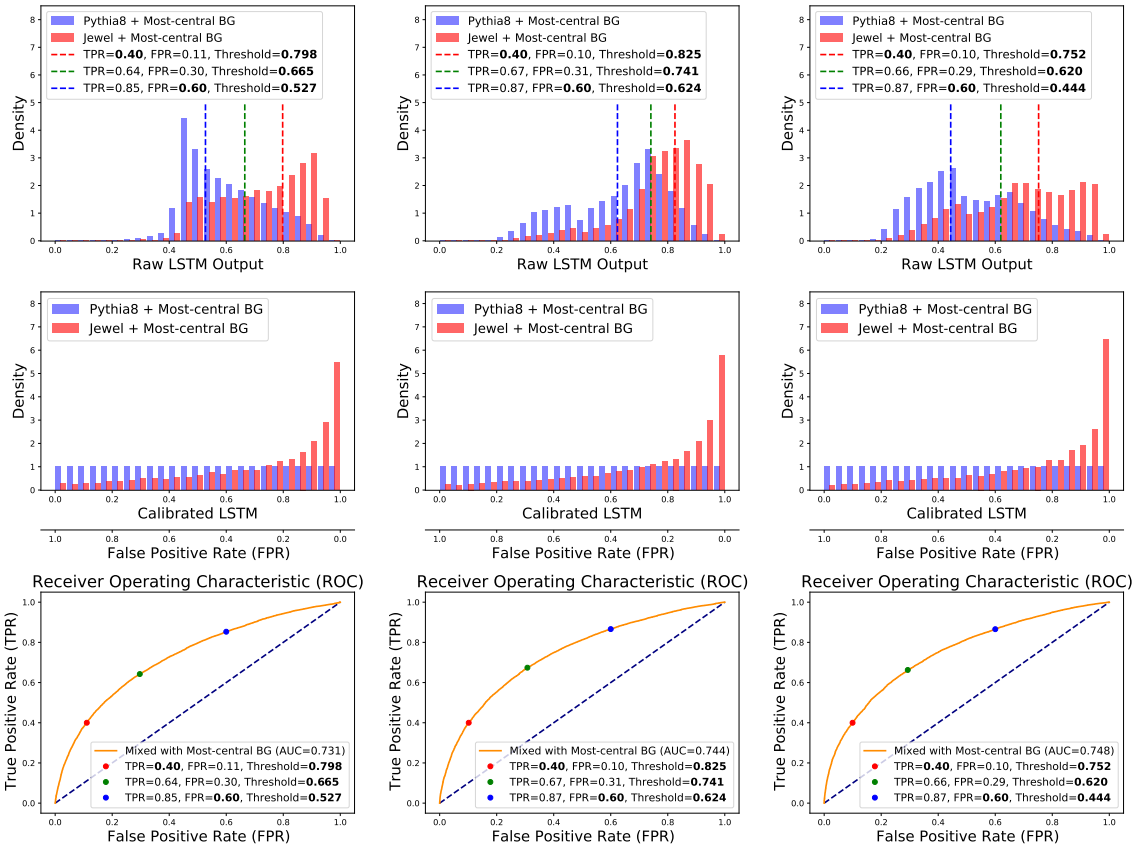

Figure 5.17: Distributions of the raw LSTM output from top 3 best trained neural networks (top), and distributions of calibrated values (mid), each with three thresholds corresponding to points on the ROC curve (bottom).

CHAPTER 6

Results

Samples of the positive class (JEWEL) are divided into two subsets, consisting of top 40% and bottom 60% samples respectively, based on raw LSTM output. To better understand the relation between LSTM output and quenching effects, we plot the distributions of substructure variables of the groomed jets (Fig. 6.1) and primary Lund plane density for the hard splittings after grooming (Fig. 6.2), for two subsets separately. In contrast, samples of the negative class (PYTHIA 8) serve as a baseline and are compared.

These two subsets of samples have shown quenching effects at different levels. The top 40% samples of the positive class show severe modification on the jet substructure with enhancement on wider and softer splittings. In comparison, the bottom 60% samples exhibit a quenching pattern that is similar to that of the negative class. Two neural networks that are trained at two different multiplicity scenarios behave consistently and show similar abilities in quantifying quenching effects, which is exciting as this machine learning approach doesn't suffer much from the density of uncorrelated underlying events. Such robustness against thermal background density may benefit from the soft drop declustering, as it exhibits high purity in locating the hard splitting for selected phase space, as studied in Ref. [73].

Our study has shown a promising machine learning approach to identifying quenching effects in the presence of uncorrelated underlying events. Sequential sub-structure variables extracted from jet clustering history carry information about the parton showering process with or without quenching effects and can be learned by an LSTM neural network. The well-trained neural network can then be used to select jets at different quenching levels. We also studied the non-deterministic behavior which happened with the supervised machine learning algorithm. With the help of the ROC curve, a calibration method is designed to avoid

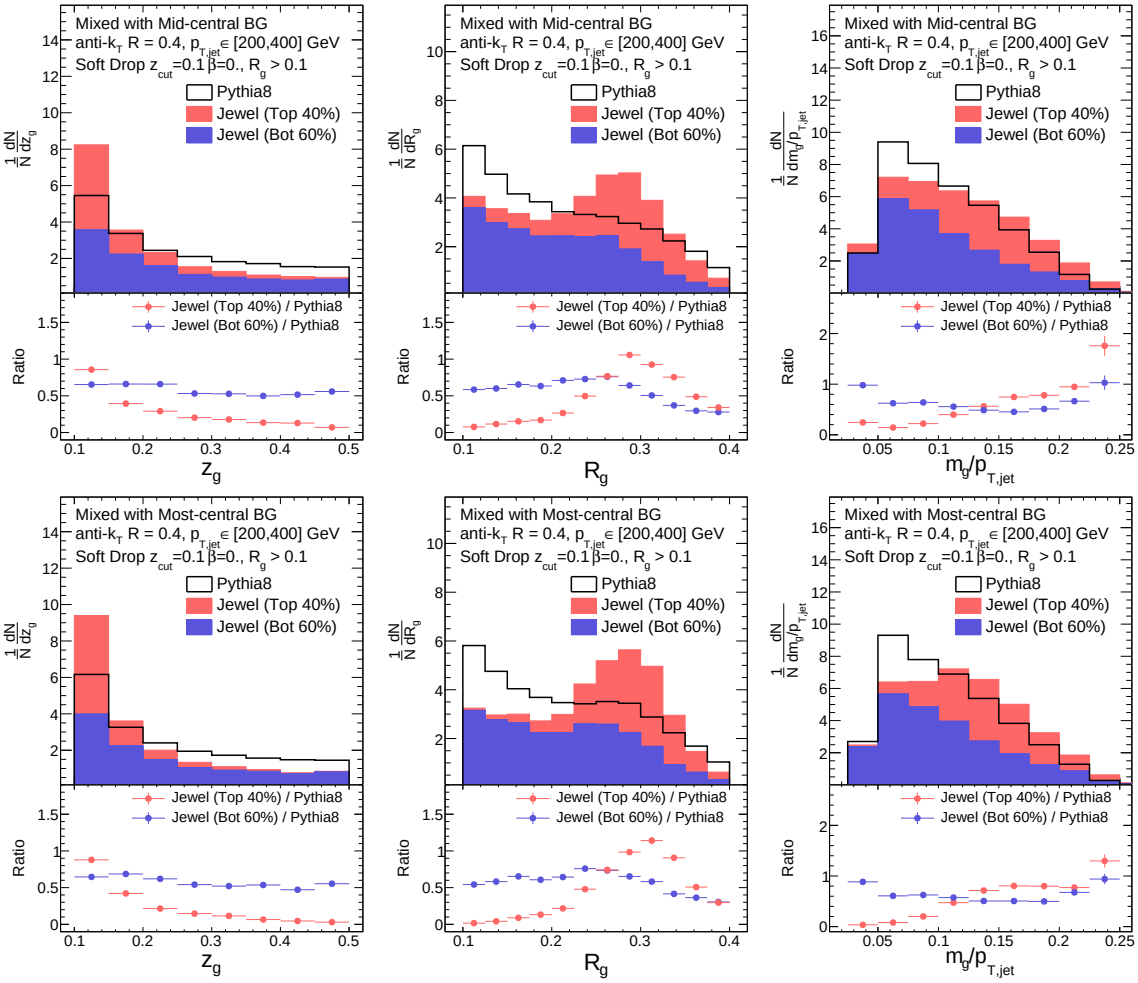

Figure 6.1: Distribution of the substructure variables of the groomed jets in the mid-central events (top) and the most-central events (bottom). Samples of the positive class (JEWEL) are plotted with stacked histograms.

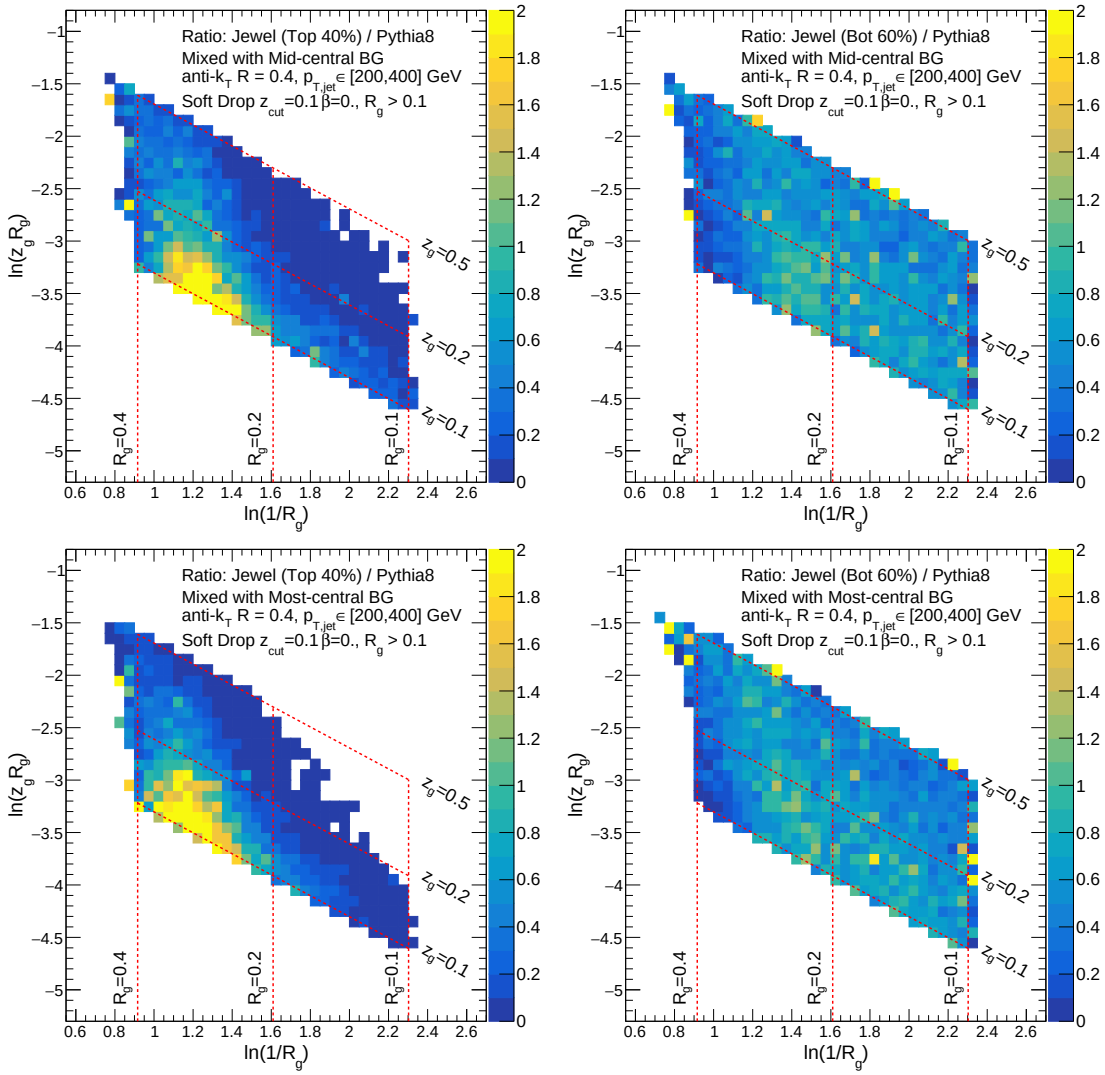

Figure 6.2: Primary Lund plane density for top 40% Jewel samples (left) and bottom 60% Jewel samples (right). Plots for the mid-central and the most-central mixing scenarios are shown on the upper and lower panels respectively.

non-determinism.

A well-trained neural network is capable of quantifying quenching effects, making it possible to divide jets simulated from a quenching model (JEWEL) into two subsets. The two separated subsets of jets have shown differences in substructure variables related to two different parton showering patterns. A vacuum-like pattern, similar to what is implemented in PYTHIA 8, and a medium-like pattern, which produces highly quenched jets.

The procedure of extracting sequential substructure variables can be used as a general feature engineering method on jets. The techniques involved in the study, such as the background subtraction and the jet grooming, have been previously applied to experimental data and their effectiveness is well-studied. Thus, our machine learning approach is very promising in the exploration of different jet classification topics with real experimental data.

Appendix A

Appendix: Sequence Padding

Considering that the sequential data in the same batch may have various lengths, thus they should be padded to equal lengths, before being processed by RNN. Typical padding methods include the pre-sequence padding and the post-sequence padding, as shown in Fig. A.1.

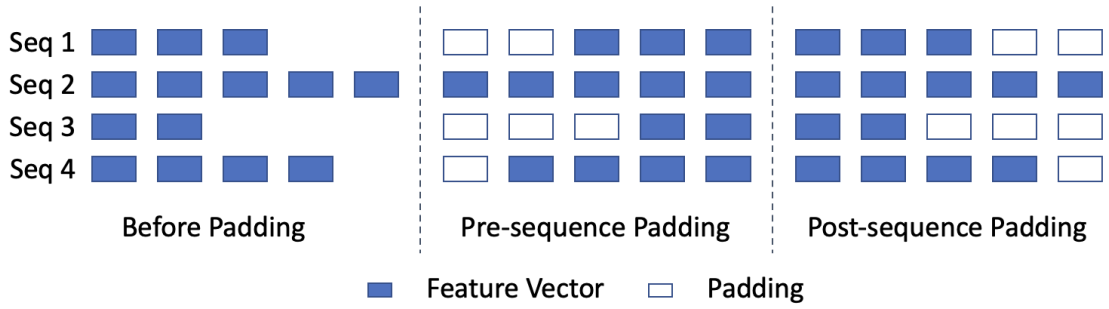

Figure A.1: Illustration of the pre-sequence and the post-sequence padding.

In this thesis, the feature vector refers to the vector x_t containing jet substructure observables at decluttering step t. Post-sequence pads of zeros are added to sequences in the same batch, such that all sequences are in equal lengths, the same as the longest sequence in the batch. The PYTORCH framework has implementations on the post-sequence padding[1].

Note that the padding procedure is applied only in the training and the validation of a neural network, during which samples are selected to form a batch. When using the trained neural network as a classifier, predictive label is made based on only one sequence, thus no padding is needed.

In addition, how padding affects the predictive labels given by a neural network is not studied. It's noticed by the author that the predictive labels can be slightly different with and without padding.

[1] Details can be found: https://pytorch.org/docs/stable/generated/torch.nn.utils.rnn.pad_sequence.html.

Appendix B

Appendix: Loss Functions

Both the mean square error (MSE) and the binary cross entropy (BCE) loss functions (Eq. 5.2) are able to guide the neural network to learn, with each having their advantages and disadvantages.

It should be pointed out that in the BCE loss function, the loss value is capped at a specific value in order to avoid infinity when x_i or y_i goes to 0 or 1, as described in the PYTORCH online maunal[1]. But the feature of BCE loss function still persists, as significant loss when x_i differs from y_i is contributed to the total training loss. As a result, fast convergence could be easily reached when training a neural network with the BCE loss function. While this feature may leads to bad noise tolerance with wrongly labeled samples. The choice of the loss function is made with compromises between training speed and noise tolerance. In this study, the MSE loss function is used.

The comparison between networks trained with BCE and MSE is vague. To achieve so, additional metrics on the performance of well-trained neural networks are required.

[1] Details can be found: https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html.

References

[1] Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, The Order of the quantum chromodynamics transition predicted by the standard model of particle physics, Nature **443**, 675 (2006), arXiv:hep-lat/0611014 .

[2] E. Annala, T. Gorda, A. Kurkela, J. Nättilä, and A. Vuorinen, Evidence for quark-matter cores in massive neutron stars, Nature Physics **16**, 907 (2020).

[3] L. R. Weih, M. Hanauske, and L. Rezzolla, Postmerger gravitational-wave signatures of phase transitions in binary mergers, Phys. Rev. Lett. **124**, 171103 (2020).

[4] A. Larkoski, S. Marzani, J. Thaler, A. Tripathee, and W. Xue, Exposing the QCD splitting function with CMS Open Data, Phys. Rev. Lett. **119**, 132003 (2017), arXiv:1704.05066 [hep-ph] .

[5] A. Tripathee, W. Xue, A. Larkoski, S. Marzani, and J. Thaler, Jet Substructure Studies with CMS Open Data, Phys. Rev. D **96**, 074003 (2017), arXiv:1704.05842 [hep-ph] .

[6] CMS Collaboration, Measurement of the Splitting Function in pp and Pb-Pb Collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, Phys. Rev. Lett. **120**, 142302 (2018), arXiv:1708.09429 [nucl-ex] .

[7] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.

[8] L. Liu, Github repository for jet quenching studies with machine learning, (2022), https://github.com/ustcllh/JetML.

[9] M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, *et al.* (Particle Data Group), Review of particle physics, Phys. Rev. D **98**, 030001 (2018).

[10] J. D. Bjorken, Highly relativistic nucleus-nucleus collisions: The central rapidity region, Phys. Rev. D **27**, 140 (1983).

[11] J. Rafelski and B. Müller, Strangeness production in the quark-gluon plasma, Phys. Rev. Lett. **56**, 2334 (1986).

[12] L. D. McLerran and T. Toimela, Photon and dilepton emission from the quark-gluon plasma: Some general considerations, Phys. Rev. D **31**, 545 (1985).

[13] T. Matsui and H. Satz, J/ψ Suppression by Quark-Gluon Plasma Formation, Phys. Lett. B **178**, 416 (1986).

[14] J. D. Bjorken, Energy Loss of Energetic Partons in Quark - Gluon Plasma: Possible Extinction of High p(t) Jets in Hadron - Hadron Collisions, (1982).

[15] M. C. Abreu *et al.* (NA50), Anomalous J / psi suppression in Pb - Pb interactions at 158 GeV/c per nucleon, Phys. Lett. B **410**, 337 (1997).

[16] E. Andersen *et al.* (WA97), Strangeness enhancement at mid-rapidity in Pb Pb collisions at 158-A-GeV/c, Phys. Lett. B **449**, 401 (1999).

[17] CERN, New state of matter created at cern, (2000), https://home.cern/news/press-release/cern/new-state-matter-created-cern, Accessed = 2022-06-12.

[18] M. Gyulassy, in *NATO Advanced Study Institute: Structure and Dynamics of Elementary Matter* (2004) pp. 159–182, arXiv:nucl-th/0403032 .

[19] BRAHMS Collaboration, Quark gluon plasma and color glass condensate at RHIC? The Perspective from the BRAHMS experiment, Nucl. Phys. A **757**, 1 (2005), arXiv:nucl-ex/0410020 .

[20] PHENIX Collaboration, Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration, Nucl. Phys. A **757**, 184 (2005), arXiv:nucl-ex/0410003 .

[21] P. Collaboration, The PHOBOS perspective on discoveries at RHIC, Nucl. Phys. A **757**, 28 (2005), arXiv:nucl-ex/0410022 .

[22] STAR Collaboration, Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions, Nucl. Phys. A **757**, 102 (2005), arXiv:nucl-ex/0501009 .

[23] P. Foka and M. A. Janik, An overview of experimental results from ultra-relativistic heavy-ion collisions at the CERN LHC: Hard probes, Rev. Phys. **1**, 172 (2016), arXiv:1702.07231 [hep-ex] .

[24] M. G. Alford, K. Rajagopal, and F. Wilczek, QCD at finite baryon density: Nucleon droplets and color superconductivity, Phys. Lett. B **422**, 247 (1998), arXiv:hep-ph/9711395 .

[25] R. Rapp, T. Schäfer, E. V. Shuryak, and M. Velkovsky, Diquark Bose condensates in high density matter and instantons, Phys. Rev. Lett. **81**, 53 (1998), arXiv:hep-ph/9711396 .

[26] S. Mukherjee, F. Rennecke, and V. V. Skokov, Analytical structure of the equation of state at finite density: Resummation versus expansion in a low energy model, Phys. Rev. D **105**, 014026 (2022), arXiv:2110.02241 [hep-ph] .

[27] F. Karsch, Lattice QCD at high temperature and density, Lect. Notes Phys. **583**, 209 (2002), arXiv:hep-lat/0106019 .

[28] STAR Collaboration, STAR Results from the RHIC Beam Energy Scan-I, Nucl. Phys. A **904-905**, 256c (2013), arXiv:1211.1350 [nucl-ex] .

[29] D. Tlusty, in *13th Conference on the Intersections of Particle and Nuclear Physics* (2018) arXiv:1810.04767 [nucl-ex] .

[30] S. Shi, *Event anisotropy v_2 at STAR*, Ph.D. thesis, Huazhong Normal University (2012).

[31] D. A. Hangal, Study of Jet Quenching in Quark Gluon Plasma Using Jet Charge Measurements, (2020), 10.25417/uic.13475553.v1.

[32] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, Glauber modeling in high energy nuclear collisions, Ann. Rev. Nucl. Part. Sci. **57**, 205 (2007), arXiv:nucl-ex/0701025 .

[33] C. Loizides, J. Nagle, and P. Steinberg, Improved version of the PHOBOS Glauber Monte Carlo, SoftwareX **1-2**, 13 (2015), arXiv:1408.2549 [nucl-ex] .

[34] D. d'Enterria, Jet quenching, Landolt-Bornstein **23**, 471 (2010), arXiv:0902.2011 [nucl-ex] .

[35] G.-Y. Qin and X.-N. Wang, Jet quenching in high-energy heavy-ion collisions, Int. J. Mod. Phys. E **24**, 1530014 (2015), arXiv:1511.00790 [hep-ph] .

[36] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigné, and D. Schiff, Radiative energy loss of high-energy quarks and gluons in a finite volume quark-gluon plasma, Nucl. Phys. B **483**, 291 (1997), arXiv:hep-ph/9607355 [hep-ph] .

[37] ALICE Collaboration, Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions, Nature Phys. **13**, 535 (2017), arXiv:1606.07424 [nucl-ex] .

[38] E. Schnedermann, J. Sollfrank, and U. W. Heinz, Thermal phenomenology of hadrons from 200-A/GeV S+S collisions, Phys. Rev. C **48**, 2462 (1993), arXiv:nucl-th/9307020 .

[39] S. A. Voloshin, A. M. Poskanzer, and R. Snellings, Collective phenomena in non-central nuclear collisions, Landolt-Bornstein **23**, 293 (2010), arXiv:0809.2949 [nucl-ex] .

[40] W. Busza, K. Rajagopal, and W. van der Schee, Heavy Ion Collisions: The Big Picture, and the Big Questions, Ann. Rev. Nucl. Part. Sci. **68**, 339 (2018), arXiv:1802.04801 [hep-ph] .

[41] ALICE Collaboration, Higher harmonic anisotropic flow measurements of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}}$=2.76 TeV, Phys. Rev. Lett. **107**, 032301 (2011), arXiv:1105.3865 [nucl-ex] .

[42] H. Niemi, K. J. Eskola, and R. Paatelainen, Event-by-event fluctuations in a perturbative QCD + saturation + hydrodynamics model: Determining QCD matter shear viscosity in ultrarelativistic heavy-ion collisions, Phys. Rev. C **93**, 024907 (2016), arXiv:1505.02677 [hep-ph] .

[43] C. Aidala *et al.* (PHENIX), Creation of quark–gluon plasma droplets with three distinct geometries, Nature Phys. **15**, 214 (2019), arXiv:1805.02973 [nucl-ex] .

[44] CMS Collaboration, Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC, JHEP **09**, 091 (2010), arXiv:1009.4122 [hep-ex] .

[45] CMS Collaboration, Multiplicity and Transverse Momentum Dependence of Two- and Four-Particle Correlations in pPb and PbPb Collisions, Phys. Lett. B **724**, 213 (2013), arXiv:1305.0609 [nucl-ex] .

[46] ATLAS Collaboration, Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV with the ATLAS detector, Phys. Lett. B **790**, 108 (2019), arXiv:1805.05635 [nucl-ex] .

[47] CMS Collaboration, Observation and studies of jet quenching in PbPb collisions at nucleon-nucleon center-of-mass energy = 2.76 TeV, Phys. Rev. C **84**, 024906 (2011), arXiv:1102.1957 [nucl-ex] .

[48] CMS Collaboration, Measurement of Jet Fragmentation in PbPb and pp Collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Phys. Rev. C **90**, 024908 (2014), arXiv:1406.0932 [nucl-ex] .

[49] CMS Collaboration, Modification of Jet Shapes in PbPb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Phys. Lett. B **730**, 243 (2014), arXiv:1310.0878 [nucl-ex] .

[50] G. Altarelli and G. Parisi, Asymptotic freedom in parton language, Nucl. Phys. B **126**, 298 (1977).

[51] CMS Collaboration, Measurement of the Splitting Function in pp and Pb-Pb Collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, Phys. Rev. Lett. **120**, 142302 (2018), arXiv:1708.09429 [nucl-ex] .

[52] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, *et al.*, Geant4—a simulation toolkit, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).

[53] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, An introduction to PYTHIA 8.2, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph] .

[54] K. C. Zapp, JEWEL 2.0.0: directions for use, Eur. Phys. J. C **74**, 2762 (2014), arXiv:1311.0048 [hep-ph] .

[55] R. Kunnawalkam Elayavalli and K. C. Zapp, Medium response in JEWEL and its impact on jet shape observables in heavy ion collisions, JHEP **07**, 141 (2017), arXiv:1707.01539 [hep-ph] .

[56] Updates to Constituent Subtraction in Heavy Ions at CMS, (2018).

[57] H. A. Andrews *et al.*, Novel tools and observables for jet physics in heavy-ion collisions, J. Phys. G **47**, 065102 (2020), arXiv:1808.03689 [hep-ph] .

[58] ALICE Collaboration, Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV, Phys. Lett. B **772**, 567 (2017), arXiv:1612.08966 [nucl-ex] .

[59] M. Cacciari and G. P. Salam, Pileup subtraction using jet areas, Phys. Lett. B **659**, 119 (2008), arXiv:0707.1378 [hep-ph] .

[60] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, Particle-level pileup subtraction for jets and jet shapes, JHEP **06**, 092 (2014), arXiv:1403.3108 [hep-ex] .

[61] P. Berta, L. Masetti, D. W. Miller, and M. Spousta, Pileup and Underlying Event Mitigation with Iterative Constituent Subtraction, JHEP **08**, 175 (2019), arXiv:1905.03470 [hep-ph] .

[62] M. Cacciari, G. P. Salam, and G. Soyez, SoftKiller, a particle-level pileup removal method, Eur. Phys. J. C **75**, 59 (2015), arXiv:1407.0408 [hep-ph] .

[63] D. Bertolini, P. Harris, M. Low, and N. Tran, Pileup Per Particle Identification, JHEP **10**, 059 (2014), arXiv:1407.6013 [hep-ph] .

[64] D. Krohn, M. D. Schwartz, M. Low, and L.-T. Wang, Jet Cleansing: Pileup Removal at High Luminosity, Phys. Rev. D **90**, 065020 (2014), arXiv:1309.4777 [hep-ph] .

[65] G. P. Salam, Towards Jetography, Eur. Phys. J. C **67**, 637 (2010), arXiv:0906.1833 [hep-ph] .

[66] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, JHEP **04**, 063 (2008), arXiv:0802.1189 [hep-ph] .

[67] M. Cacciari, G. P. Salam, and G. Soyez, FastJet User Manual, Eur. Phys. J. C **72**, 1896 (2012), arXiv:1111.6097 [hep-ph] .

[68] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Jet substructure as a new Higgs search channel at the LHC, Phys. Rev. Lett. **100**, 242001 (2008), arXiv:0802.2470 [hep-ph] .

[69] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, JHEP **02**, 084 (2010), arXiv:0912.1342 [hep-ph] .

[70] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches, Phys. Rev. D **81**, 094023 (2010), arXiv:0912.0033 [hep-ph] .

[71] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, JHEP **05**, 146 (2014), arXiv:1402.2657 [hep-ph] .

[72] S. Carrazza and F. A. Dreyer, Jet grooming through reinforcement learning, Phys. Rev. D **100**, 014014 (2019), arXiv:1903.09644 [hep-ph] .

[73] J. Mulligan and M. Ploskon, Identifying groomed jet splittings in heavy-ion collisions, Phys. Rev. C **102**, 044913 (2020), arXiv:2006.01812 [hep-ph] .

[74] F. A. Dreyer, G. P. Salam, and G. Soyez, The Lund Jet Plane, JHEP **12**, 064 (2018), arXiv:1807.04758 [hep-ph] .

[75] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images — deep learning edition, JHEP **07**, 069 (2016), arXiv:1511.05190 [hep-ph] .

[76] J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman, Jet-Images: Computer Vision Inspired Techniques for Jet Tagging, JHEP **02**, 118 (2015), arXiv:1407.5675 [hep-ph] .

[77] *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*, Tech. Rep. (CERN, Geneva, 2017) all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-017.

[78] A. Sherstinsky, Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network, Physica D: Nonlinear Phenomena **404**, 132306 (2020).

[79] S. Hochreiter and J. Schmidhuber, Long short-term memory, Neural computation **9**, 1735 (1997).

[80] F. A. Dreyer, L. Necib, G. Soyez, and J. Thaler, Recursive Soft Drop, JHEP **06**, 093 (2018), arXiv:1804.03657 [hep-ph] .

[81] L. Apolinário, N. F. Castro, M. Crispim Romão, J. G. Milhano, R. Pedro, and F. C. R. Peres, Deep Learning for the classification of quenched jets, JHEP **11**, 219 (2021), arXiv:2106.08869 [hep-ph] .

[82] J. Bergstra, D. Yamins, and D. Cox, in *Proceedings of the 30th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 28, edited by S. Dasgupta and D. McAllester (PMLR, Atlanta, Georgia, USA, 2013) pp. 115–123.