UNIVERSITY OF CALIFORNIA

Los Angeles

Statistical Aspects of Jets Containing Heavy Quarks

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Physics

by

Jared Denali Reiten

2023

ABSTRACT OF THE DISSERTATION

Statistical Aspects of Jets
Containing Heavy Quarks

by

Jared Denali Reiten
Doctor of Philosophy in Physics
University of California, Los Angeles, 2023
Professor Zhongbo Kang, Chair

Jets comprise a rich class of emergent phenomena stemming the underlying theory of the strong nuclear force, Quantum Chromodynamics. As jets are produced in copious quantities in hadron colliders, understanding their internal structure and evolution is of the utmost importance for modern particle physics. In this thesis, we study various aspects of a special class of jets—that is, jets containing heavy quarks, such as charm, bottom and top—which can all be understood from a statistical point of view. In the first part, we consider situations in which the observation of back-to-back heavy-quark dijet pairs shed light on key physics governing the final and initial states of high-energy particle collisions—from the modification of dijet mass spectra by the quark-gluon plasma created in the collisions of heavy ions to the probing of the Sivers spin asymmetry in deep inelastic scattering. In the second part, we analyze the internal landscapes of jets initiated by heavy quarks and demonstrate how the so-called "dead-cone" effect manifests in the cumulants of jet substructure distributions. In the third and final part, we adapt concepts from the machine learning community to tag top jets from a background of jets initiated by light quarks and gluons as well develop a novel data type that is particularly well-suited to exposing the characteristic angular structure of top decay products.

The dissertation of Jared Denali Reiten is approved.

E. T. Tomboulis

Huan Z. Huang

Zvi Bern

Zhongbo Kang, Committee Chair

University of California, Los Angeles

2023

To my mother

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Firstly, I would like to thank my PhD advisor, Zhong-Bo Kang for all the support and guidance over the years. I have learned a tremendous deal from you in terms of physical insight, career navigation, and leadership.

I would like to thank my mentors Andrey Sadofyev and Andrew Larkoski for the countless hours spent in conversation, at the chalkboard, and nose-deep in Mathematica notebooks with me. A great deal of my scientific style may trace its lineage through the both of you.

I would like to thank Maarten Buffing, Ding-Yu Shao, Farid Salazar, Wai-Kin Lai, Yiyu Zhou, Jani Penttala, Kyle Lee, John Terry, Fanyi Zhao, and Sasza Czajka for all the enjoyable times spent in and out of the office. I am grateful for the many laughs and pieces of wisdom you have all shared with me.

I thank Tejes Gaertner for all our time spent thinking up projects, carrying out calculations, and writing up papers. I am particularly thankful for your willingness to discuss physics at any hour of the day on any day of the week. It has been a true pleasure to see you grow as a scientist and I am proud to consider you a peer.

I would like to thank Stacy Kravetz and Amy Levin for sharing their memories of Jay Goldberg with me and for teaching me what it means to be a mensch. Your family is forever in my heart.

I thank John Fahey and George Ciampa for their service to this nation and for everything they taught me about life. I am deeply honored to have been able to sit at the same tables as you.

I thank my friends Colin Steinbroner, Thomas Gross, and Nathan Hendel for all the years spent in laughter, hardship, and victory. You are a fine group of men.

I thank all the Reitens that came before me, in particular, Thomas, Julie, and Dan. “We will see you when we see you.”

Above all, I thank my mother, Barbara. Neither English nor mathematics can express the love and gratitude I have for you. All good things that have and are to come of my life are truly due to you—only the mistakes are mine.

VITA

2011–2015 B.A. Physics and Applied Mathematics, University of California, Berkeley.

2018 M.S. Physics, University of California, Los Angeles.

2017–2019 Teaching Assistant, Department of Physics and Astronomy, University of California, Los Angeles.

2019–2021 UC-National Lab In-Residence Graduate Fellow, University of California, Los Angeles and Los Alamos National Laboratory.

2021–2023 Teaching Assistant, Department of Physics and Astronomy, University of California, Los Angeles.

PUBLICATIONS

1. "An infrared flavor for heavy-quark jets"
   Z. B. Kang, A. J. Larkoski and J. Reiten, [arXiv: 2401.xxxxx [hep-ph]]

2. "The dead-cone angle through cumulants of subjet-energy functions"
   Z. B. Kang, A. J. Larkoski and J. Reiten, [arXiv: 2401.xxxxx [hep-ph]]

3. "Unsupervised learning in the metric space of jets"
   T. Gaertner, Z. B. Kang and J. Reiten, [arXiv: 2312.06948 [hep-ph]]

4. "The simplicial substructure of jets"
   T. Gaertner and J. Reiten, [arXiv: 2312.06596 [hep-ph]]

5. "Jet quenching in anisotropic flowing matter"
   M. V. Kuzmin, X. M. Lópex, J. Reiten and A. V. Sadofyev, [arXiv: 2309.00683 [hep-ph]]

6. "Resummation of the Sivers asymmetry in heavy flavor dijet production at the Electron-Ion Collider"

   Z. B. Kang, J. Reiten, D. Y. Shao and J. Terry, *Proceedings of **DIS2021**: XXX International Workshop on Deep-Inelastic Scattering and Related Subjects, Stony Brook University, New York, USA, 12-16 April 2021*, [arXiv: 2108.07257 [hep-ph]]

7. "Magnetic monopole in a chiral plasma: chiral dyon"

   M. Lublinsky, J. Reiten and A. V. Sadofyev, Phys. Rev. D **103** (2021) 12, 125017, [arXiv: 2108.07257 [hep-ph]]

8. "QCD evolution of the gluon Sivers function in heavy flavor dijet production at the Electron-Ion Collider"

   Z. B. Kang, K. Lee, J. Reiten, D. Y. Shao and J. Terry, JHEP **05** (2021) 286, [arXiv: 2012.01756 [hep-ph]]

9. "The dijet mass distribution in heavy ion collisions"

   J. Reiten, *Proceedings of **ISMD2019**: XLIX International Symposium on Multi-Particle Dynamics, Santa Fe, New Mexico, USA, 9-13 September 2019*, [arXiv: 2002.12442 [hep-ph]]

10. "Drag force to all orders in gradients"

    J. Reiten and A. V. Sadofyev, JHEP **07** (2020) 146 , [arXiv: 1912.08816 [hep-th]]

11. "Light and heavy flavor dijet production and dijet mass modification in heavy ion collisions"

    Z. B. Kang, J. Reiten, I. Vitev and B. Yoon, Phys. Rev. D **99** (2019) 3, 034006, [arXiv: 1810.10007 [hep-ph]]

12. "Thermal conductivity measurements of PTFE and Al_2O_3 ceramic at sub-Kelvin temperatures"

    A. Drobizhev, J. Reiten, V. Singh and Y. G. Kolomensky, Cryogenics **85** (2017) 63-70, [arXiv: 1612.02514 [cond-mat.mtrl-sci]]

CHAPTER 1

Introduction

In this chapter, we cover some basic preliminaries underpinning the work of this thesis. First, we provide the fundamental motivation for the study of jets containing heavy quarks, as outlined by the Nuclear Science Advisory Committee Long Range Plan for Nuclear Science. Second, we provide an introduction to Quantum Chromodynamics as well as a basic computation for the production of a jet of radius R. We finish with a summary detailing the structure of this thesis.

1.1 Motivation

In order to probe the excitations of spacetime at the smallest possible length scales, one needs to collide hadrons at the highest achievable energies. Such is the goal of collider physics and in attaining this goal, the scientific community has amassed a thorough and precise knowledge of the Standard Model (SM) of particle physics. The SM describes three of the four fundamental interactions occurring in nature, that is, the electromagnetic, weak, and strong interactions. Of these, the sector responsible for the strong nuclear force is known as Quantum Chromodynamics (QCD), which plays an incredibly significant role in collider physics. QCD describes the interactions of elementary particles known as quarks and gluons, but due to the most famous property of QCD—that is, asymptotic freedom [GW73, Pol73]—these degrees of freedom can only be probed through the exchange of exceedingly large amounts of energy. The resulting high-energy excitations then lead to radiative cascades of yet even more quarks and gluons which are produced in collimated

beams leaving the collision site. As their energies lower, the quarks and gluon coalesce to form hadrons, and these clusters of hadrons, known as jets, are what ultimately reach particle detectors. The study of jet physics is central to collider physics, as jets are the most highly-produced product of colliders.

A frontier of great interest is the study of jets initiated by heavy quarks, such as charm and bottom, as such jets provide access to important quantities characterizing the initial and final states of nuclear collisions. Consider the case of heavy ion collisions, in which the collision of nuclei result in the formation of the hot and dense liquid known as the quark-gluon plasma (QGP). The energy loss experienced by quarks as they move through this fluid is well-known to depend strongly on the flavor of the quark, as physical attributes such as the quark mass have an interesting interplay with the physics of the medium. Bottom quarks, which are about one thousand times heavier than light quarks, produce unique energy loss signatures due to their large mass, $m_b \simeq 4.2$ GeV [DGV06]. Jets containing at least one bottom quark inside them are called bottom jets, or b-jets in short. Due of their large mass, bottom quarks are produced rarely in heavy ion collisions, thus b-jets constitute but a small fraction of the total jets produced. When b-jets traverse the QGP created in the same collisions, they interact strongly with the medium. Consequently, their production rate as well as the resulting internal particle distribution of the jet (so-called jet substructure) is modified due to the existence of the QGP, relative to the vacuum case. By studying the pattern of these modifications, one can infer valuable information about the properties of the QGP.

One of the central goals of the sPHENIX experiment at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory is to provide state-of-the-art capabilities for studies of the strongly interacting QGP through the use jet and heavy-flavor observables, of which b-jets are one of the golden channels. In this regard, all work dedicated to the generation of predictions for b-jet observables is extremely timely and is a direct response to the urgent need for theoretical progress along this direction. The primary motivation for the work of this thesis is to provide such predictions in order to make an immediate/critical impact on the sPHENIX experimental program. The success of sPHENIX is of the utmost importance to the scientific community, as outlined in the Long Range Plan (LRP) for Nuclear Science in 2015 by the Nuclear Science Advisory Committee (NSAC) [Apr15]. Crucially, this aim was recently reiterated and deemed "essential"

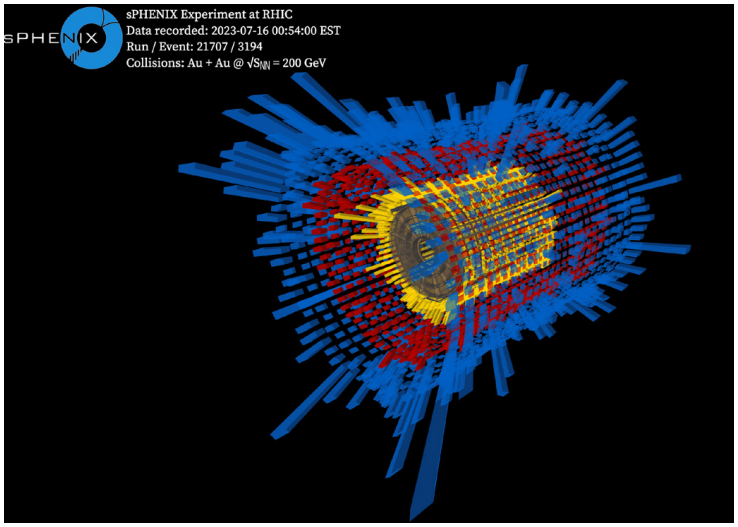

Figure 1.1: Event display from an Au+Au collision at RHIC captured by the sPHENIX detector, obtained from 2023 NSAC LRP [Com23]. Such a visualization highlights the overwhelming number of jets in the produced in the final states of such collisions.

in the NSAC LRP for 2023 [Com23]. Thus the need for detailed theoretical predictions is of the utmost importance.

The key first step towards progress in this direction is the analytic calculation for jet substructure observables for the vacuum baseline for sPHENIX. A glance at Fig. 1.1 is enough to convince the reader that the final states produced by heavy-ion collisions in sPHENIX are of extremely high complexity. In order to characterize the modification of heavy-flavor substructure observables by the medium, one must have the cleaner vacuum case under theoretical control—and this is non-trivial task, due to the fact that b-jet production involves several widely-separated energy scales, such as its mass m_b, the transverse momentum of the jet p_T^{jet}, as well as a dynamic scale related to the size of the jet's local domain. For the case of jet substructure, yet additional scales exist due to the measurement of the internal structure. As a consequence, computations involving b-jets are complicated and subtle. The usual fixed-order perturbative calculations in the expansion of strong coupling constant are impaired by the existence of logarithms of the ratios of the various aforementioned scales of the problem. These terms persist in all orders of the perturbative series and one must sum these contributions to all orders to be able to render convergence of the calculation.

On the topic of the substructure of jets containing heavy quarks, the ALICE Collaboration

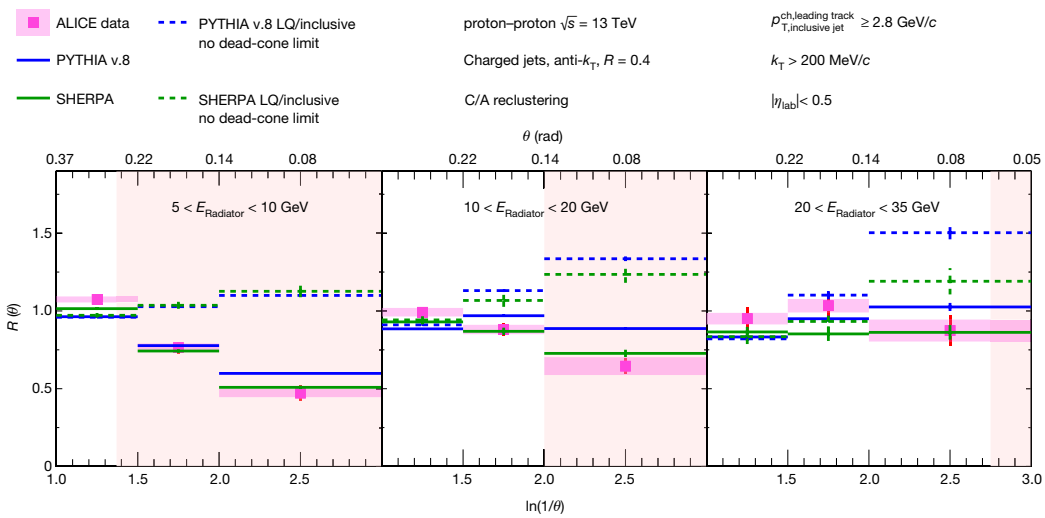

Figure 1.2: Demonstration of the dead-cone effect, as captured by the measurement of the $R(\theta)$ observable by the ALICE Collaboration [Ach22] at CERN.

at the Large Hadron Collider (LHC) at CERN performed a remarkable measurement of what is known as the "dead-cone" effect for charm-initiated jets [Ach22]. The dead-cone effect refers to the suppression of gluon radiation off a heavy quark within an conical region of angular extent $\theta_{\mathrm{dc}} \simeq m_Q/E_Q$ where m_Q and E_Q are the mass and energy, respectively, of the heavy quark under consideration—θ_{dc} is referred to as the dead-cone angle. Such a measurement was a scientific feat as the dead-cone angle is one of the principal ways in which the heavy-quark mass imprints itself on the internal radiation pattern of a jet. However, the precise definition of the observable used by the ALICE Collaboration to measure this effect—in particular the requisite sequence of jet-reclustering steps that must be experimentally performed—do not lend themselves to theoretical calculation within the framework perturbative QCD (pQCD). As such, it is important for theorists to come up with complementary experimental observables which facilitate theoretical prediction. Doing so works to maintain the close feedback loop between theory and experiment that has lead to QCD being one of the most successful and fruitful branches of physics over the past fifty years [Gro22]. Therefore, the urgent need for theoretical progress in the computation of heavy-quark jet substructure observables as mentioned in the context of the sPEHNIX experiment at RHIC is echoed by the experimental program at the LHC as well.

Lastly, as the LHC achieves the highest center-of-mass energies of any modern collider, it pro-

vides access to jets containing the heaviest of all quarks, that is, the top quark. Of course, one should be careful with their language in this case, as while the LHC certainly produces jets *initiated* by top quarks—the large mass and subsequent decay of the top makes it so that only its decay products at the collinear radiation stemming therefrom are truly "contained" in the final jet itself. However, just as the mass of the charm and bottom imprint themselves on the final radiation pattern of their jets through the dead-cone angle, the top mass leaves its mark through the characteristic angular spread of its decay products. This gives jets initiated by top quarks a distinctive substructure relative to those initiated by light quarks and gluons—typically referred to simply as "QCD jets." Thus, the task of top-tagging can be achieved through the intensive analyses of QCD and top jet profiles, where the QCD jets form the "background" for the top jet "signal." Such a task lends itself naturally to the application of advanced Machine Learning (ML), Deep Neural Network (DNN), and general Artificial Intelligence (AI) techniques. While the particle physics community has certainly benefitted from such applications, many of the architectures employed rely on tens of thousands (or more) parameters whose sheer number obfuscates any physical understanding for the efficacies of such techniques. In recent years, the notion of "explainable AI" has taken a more prominent role in guiding the vision of AI applications in particle physics [NR22], and the reason for this is rather intuitive for physicists to grasp. The goal of physics is not simply taking a detailed account of the world around us, but rather to *explain* and *understand* the world.

The aim of this thesis is to address all of these points. In it, we define novel observables consisting of heavy-quark jets that shed light on important questions regarding heavy flavor of relevance for the sPEHNIX experiment. Furthermore, we perform analytic predictions for charm- and bottom-initiated jets that reveal the dead-cone effect in a theoretically-calculable fashion. Lastly, we develop novel techniques for the tagging of top jets that can be understood using only but a few parameters whose physical origins are clear.

1.2 Structure of thesis

This thesis follows a structure that can be divided into three parts, with each part containing a pair of closely-related themes and physical goals.

Part I pertains to observables constructed through the observation of pairs of jets containing heavy quarks, known as dijets, created high-energy particle collisions in a back-to-back configuration. Such configurations provide access to particular flavor channels in the underlying hard collisions, and thus allow for the probing of specific features of the initial/final partonic states thereof. Ch. 2 concerns itself primarily with the mass spectrum of dijets emerging from the collision of heavy ions and their subsequent modification pattern induced by their traversal of the quark-gluon plasma. Ch. 3 shows how heavy-flavor dijets can be leveraged in order to extract the gluonic contribution to the Sivers asymmetry of the proton.

In Part II, we venture into the internal substructure of jets initiated by charm and bottom quarks to study what is known as the dead-cone effect. This is the principal way in which the heavy quark mass m_Q imprints itself onto the radiation pattern in the immediate vicinity of heavy quarks. Ch. 4 considers the fragmentation of a jet into a smaller subjet centered about a jet axis which preferentially aligns itself with the trajectory of the initiating heavy quark. Cumulants in the distribution of subjet energy-fractions are shown to capture the dead-cone effect in a striking fashion. In Ch. 5 we modify the previous setup to (1) provide a reliable definition of a jet's flavor with respect to partonic degrees of freedom, and most importantly, their masses, (2) to compute the posterior probabilities for particular hard-scattering channels to give rise to heavy-quark jets through Bayesian inference, and (3) to provide a distribution whose cumulants not only reveal the dead-cone effect, but does so in a way that is nearly universal across collider kinematics.

Part III focuses on illustrating the way in which jets initiated by the decay of top quarks differ from those initiated by massless quarks and gluons through various geometries which their basic data give rise to. In Ch. 6 we demonstrate how the geometry induced by an Optimal Transport metric can be leveraged for the unsupervised clustering/tagging of top jets from those due to light degrees of freedom—and in a way that makes use of a minimal number of parameters whose optimal values can be intuited physically. Ch. 7 then presents a new data type for the representation of jets, whose resultant geometric and information-theoretic structures provide access to the top mass m_{top} though a canonical angular resolution scale.

The overarching goal of this thesis is to provide a family of techniques that allows the community to progress in understanding the dynamics of heavy quarks and the development of their

highly non-trivial radiative cascades as they evolve into jets. While analyses are performed through a wide range of means—including the use of Monte Carlo simulations, analytic and quantum-field-theoretic calculations, as well as the application of machine learning algorithms—all methods can be understood to fall under the framework of statistics. We hope that both the power of such a framing's robustness as well as the elegance of its simplicity can be appreciated.

1.3 Jets and Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the quantum filed theory describing the strong nuclear force—the force responsible for the binding of hadrons and the stability of atomic nuclei. QCD recently celebrated its fiftieth anniversary [Gro22], and has seen tremendous success in the decades following its inception. The fundamental degrees of freedom are the elementary particles known as quarks and gluons. The quarks are represented by Dirac spinor fields while the gluons correspond to non-Abelian gauge fields. As such, the Lagrangian for QCD is remarkably simple to state [Sch14]:

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4}\left(F^a_{\mu\nu}\right)^2 + \sum_f \bar{\psi}^i_f \left[i\gamma^\mu \left(\delta_{ij}\,\partial_\mu + g_s\, A^a_\mu\, t^a_{ij}\right) - m_f\,\delta_{ij} \right] \psi^j_f \,, \tag{1.1}$$

where sum runs over quark flavors $f \in \{u, d, s, c, b, t\}$, i, j denote Dirac indices, μ is a Lorentz index, and a the color index—repeated instances of μ and a are understood to be summed over. γ^μ form the usual set of Dirac matrices and m_f is the bare mass of the quark field of flavor f. ψ denotes the spin-$1/2$ matter field of the quarks with A being the gauge field whose Yang-Mills field strength tensor is given by

$$F^a_{\mu\nu} \equiv \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu \,. \tag{1.2}$$

The group under which QCD is gauged is $\mathrm{SU}(3)$, whose $N_c^2 - 1 = 8$ generators t^a satisfy the algebra

$$\left[t^a, t^b\right] = i f^{abc}\, t^c \,, \tag{1.3}$$

where f^{abc} are the structure constants.

Perhaps the most pivotal feature of QCD is the running of the strong coupling constant, which governs the interaction strength between gluons with themselves as well as with quarks. The combination which characterizes such interactions is referred to as α_s, which is given by

$$\alpha_s \equiv \frac{g_s^2}{4\pi}\,, \tag{1.4}$$

in direct analogy with the fine-structure constant specifying the interaction strength of the electromagnetic force. The running of α_s is the following statement:

$$\mu^2 \frac{\partial}{\partial \mu^2} \alpha_s(\mu^2) = \beta(\alpha_s)\,, \tag{1.5}$$

where $\beta(\alpha_s)$ is referred to simply as the "β-function," which can be interpreted as an effective velocity field governing the dynamical evolution of α_s, with $\log \mu^2$ playing the role of a time variable. The β-function itself can be expressed as a power series in the strong coupling α_s [ESW11]

$$\beta(\alpha_s) = -\alpha_s \sum_{n=0}^{\infty} \beta_n \left(\frac{\alpha_s}{4\pi}\right)^{n+1}\,, \tag{1.6}$$

where the first few coefficients are

$$\begin{aligned}
\beta_0 &= 11 - \frac{2}{3} n_f\,, \\
\beta_1 &= 102 - \frac{38}{3} n_f\,, \\
\beta_2 &= \frac{2857}{2} - \frac{5033}{18} n_f + \frac{325}{54} n_f^2\,,
\end{aligned} \tag{1.7}$$

with n_f denoting the number of active quark flavors. Let us consider solely the leading order (LO) term appearing in Eq. (1.6) and consider the case of five active quark flavors (hence set $n_f = 5$):

$$\beta^{(\mathrm{LO})}(\alpha_s) = -\frac{23}{12\pi} \alpha_s^2\,, \tag{1.8}$$

and hence is manifestly negative. This is to say that as we increase the energy resolution scale of an experiment μ, the effective value of α_s *decreases* and vice versa. This is the famous demonstration of asymptotic freedom, which provided unequivocal evidence for QCD as the theory of the strong interaction, and for which Gross, Politzer, and Wilczek were awarded the Nobel Prize in Physics—see [GW73, Pol73] for their monumental calculations. What asymptotic freedom teaches us about nature is that as we probe nuclear structures, e.g. a proton, at high energies, we resolve the point

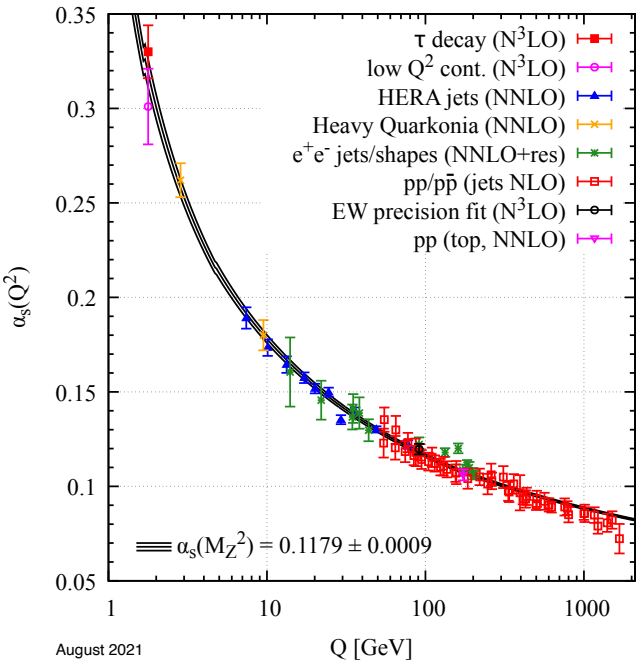

Figure 1.3: Collection of measurements of the strong coupling constant α_s as a function of energy resolution scale Q, compiled by the particle Data Group (PDG) [Zyl20]. The legend delineates the processes from which the extraction is made as well as the fixed perturbative order to which the hard-scattering cross section is computed.

particles which comprise the proton—the quarks and gluons—propagating as nearly-free entities. An incredibly consequential corollary to this is that as we resolve the proton etc. at lower energies, these once-free constituents become tightly-bound and form the macroscopic color-neutral particle that is the proton itself. This is the onset of confinement, whose scale is characterized by $\Lambda_{\mathrm{QCD}} \simeq 250$ MeV.

Solving Eq. (1.5) keeping only the zeroth-order term in the expansion of $\beta(\alpha_s)$ we obtain the following relation

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{\alpha_s(\mu^2)}{4\pi}\beta_0 \log \frac{Q^2}{\mu^2}}\,, \tag{1.9}$$

which gives α_s as measured at a scale Q in terms of α_s as observed at the reference scale μ, effectively allowing us to freely translate the coupling between different levels of resolution. Again, the above relation is only the LO expression—the inclusion of higher-order corrections can be visualized in Fig. 1.3, where we see that the agreement between theory and experiment holds to a

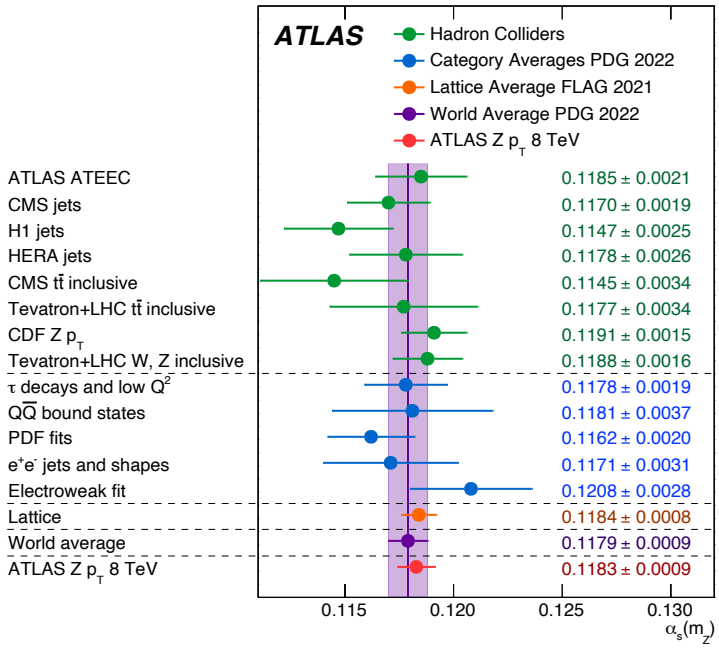

Figure 1.4: Summary of $\alpha_s(M_Z)$ extractions from various processes together with the measurement performed by the ATLAS Collaboration [Aad23] at CERN. This extraction holds the record for precision among all experimental determinations.

remarkable level of precision. It is important to comment that the standard reference scale with respect to which values of α_s are often quoted is that of the Z^0 mass, $M_Z \simeq 91.2$ GeV. Quite recently, the ATLAS Collaboration at CERN reported the highest-precision experimental determination of α_s to date, with $\alpha_s(M_Z) = 0.1183 \pm 0.0009$—see Fig. 1.4 for this measurement and its comparison to other leading determinations. Such a measurement is not only a remarkable feat of experimental ingenuity but is also a fitting homage to nature as we celebrate fifty years of QCD.

Hopefully our discussion thus far has demonstrated that behind the simplicity of Eq. (1.1) live incredibly rich physical phenomena with far-reaching consequences for the structure of matter. This thesis concerns itself primarily with the emergent phenomena known as jets, that is, collimated sprays of hadrons which emerge from the collisions of hadrons at ultra-relativistic speeds. While the dynamics of jet evolution are intricately tied in with the running of the coupling α_s, another key feature of QCD is required to lay the groundwork. As QCD consists of massless particles, it is a theory in which we must understand and wrangle the infrared divergences which invariably

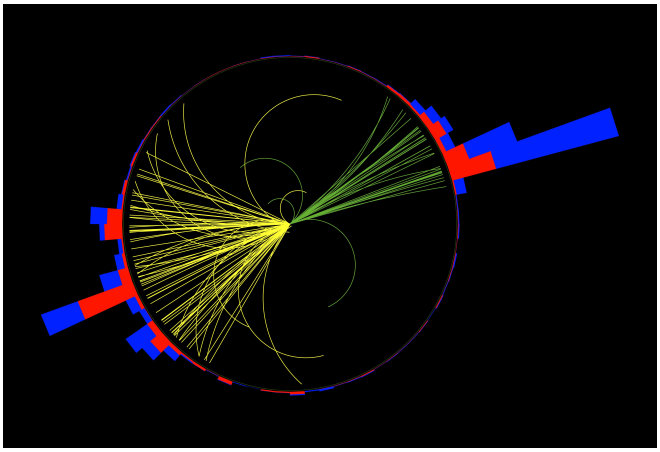

Figure 1.5: Event display from the LHC depicting back-to-back jet production, provided by the CMS Collaboration [Col20].

come [Wei, Sch14].

To keep our discussion simple, what infrared divergences signal is the enhanced probability for a particle to emit an uncontrolled number of extremely low-energy particles along the direction collinear to the motion of the emitter. With proper care and summing over all virtual and real emission diagrams that appear at any fixed order in α_s, one may "tame" such divergences and achieve a finite result. Sterman and Weinberg went further to help us understand the physical nature of such taming, and in the process developed the first theoretical formalism for jets [SW77].

We may understand jets intuitively as follows. We begin with a quark or gluon emerging from a collision at high virtuality. This virtuality is decreased through successive emissions of soft and collinear particles, where with each emission, a fraction of the initiating particle's energy is redistributed to those emitted. This process continues until the energy distributed across the slew of particles approaches a low resolution scale, say of $\sim \mathcal{O}(1)\,\mathrm{GeV}$ at which point α_s begins to ramp up and confinement sets it. The final state is then a large collection of hardons conically-centered about the trajectory of the initiating quark or gluon. Such final states can be visualized beautifully and event display of the the CMS Collaboration pictured in Fig. 1.5.

A deeply interesting fact about jets, which is the central avenue of exploration of this thesis, is that the final radiation pattern within a jet depends non-trivially of the flavor of the flavor of the initiating parton. This can be visualized in Fig. 1.6 which depicts images of jets initiated by

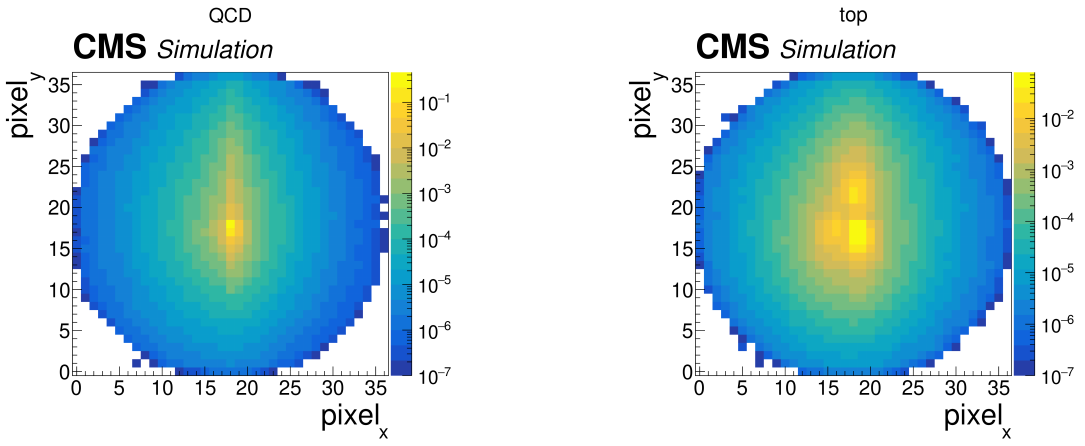

Figure 1.6: Examples of averaged jet images originating from different particles. The left depicts that due to light quarks and gluons (i.e. "QCD" jets) while the right is for jets initiated by top quarks. Such a visualization provides intuition for how jets "appear" to the calorimeter cells which detect them. Image provided by the CMS Collaboration [Sir20].

light quarks and gluons as well as by top quarks, the heaviest-known elementary particle. In these images, the pixel intensity is a measure of the relative energy concentrated at a particular location. These images are averages over an ensemble of jets of each class, as viewed by the calorimeter cells onto which they deposit themselves. As we see, jets initiated by light quarks and gluons, labeled as "QCD" jets in Fig. 1.6, are highly-collimated on average showing that they predominantly exhibit a single core of high energy flow. This is to be sharply contrasted with the average top jet image, which displays not only extended structure in its distribution of energies, but also the existence of more than one core of dominant energy flow—a relic of the detailed decay structure undergone by top quarks almost immediately after their inception. Thus, jets provide us with an interesting lens through which we may peer into the workings of nature at high energies, or equivalently the excitations of spacetime at the smallest of achievable scales—that is, through the complex multi-particle state collected by calorimeter cells, imprints of the fundamental quark and gluon degrees of freedom contained in the QCD Lagrangian of Eq. (1.1) are contained. This remarkable fact underpins all the work contained in this thesis.

Before concluding this introduction, it would be highly-useful to demonstrate what goes into computing the rate of jet production for a simple physical process. Doing so will hopefully not only

provide intuition for the more detailed computations to come, but also breathe life into Eq. (1.1).

Jets comprise the final state of any collisional process that couples to QCD. Perhaps the simplest of all such processes is electron-positron scattering, for in such collisions QCD is involved purely in the final state—the initial being completely determined by Quantum Electrodynamics (QED). What is better yet is the historical significance that such collisions have played in the determination of QCD as the theory of strong nuclear force. By now, such calculations are completely standard and we follow those laid out in [Sch14].

First, let us consider the process, governed at LO purely by QED, of electron-positron scattering into a muon-anti-muon final state. This process has cross section

$$\sigma(e^+e^- \to \mu^+\mu^-) = \frac{4\pi\alpha_{\rm EM}^2}{3E_{\rm CM}^2} \equiv \sigma_0 \,. \tag{1.10}$$

Instead of muons in the final state, we can instead couple to QCD via the electromagnetic interaction with a quark-anti-quark pair of flavor f:

$$\sigma(e^+e^- \to \bar{q}_f q_f) = \sum_{\rm colors} \sigma_0 \, Q_f^2 = N_c \, \sigma_0 \, Q_f^2 \,, \tag{1.11}$$

where we see that all is identical to Eq. (1.10) except for the requisite sum over colors and the effective modification of $\alpha_{\rm EM}$ due to the fractional electric charges of the quarks. Now, we may be extremely inclusive with our final state, which is to say, by being exceedingly general with what we measure, we may analytically compute a non-trivial hadronic final state to LO. The inclusive final state we are alluding to is simply that of hadrons, with no kinematic cuts or any further measurements of their properties. To obtain such a cross section we may simply sum Eq. (1.11) over quark flavors (up through b for the center-of-mass-energies we consider in what follows),

$$\sigma(e^+e^- \to \text{hadrons}) = \sum_{f=\{u,d,\ldots,b\}} \sigma(e^+e^- \to \bar{q}_f q_f) \,. \tag{1.12}$$

Now, we may take the ratio of this cross section to that for the muonic final state to cancel the like-terms corresponding to the initial electron-positron state and thus purely isolate the number of quark colors in QCD:

$$R \equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} = \sum_{\rm colors} \sum_{f=\{u,d,\ldots,b\}} Q_f^2 = N_c \sum_{f=\{u,d,\ldots,b\}} Q_f^2 \simeq 3.67 \,, \tag{1.13}$$

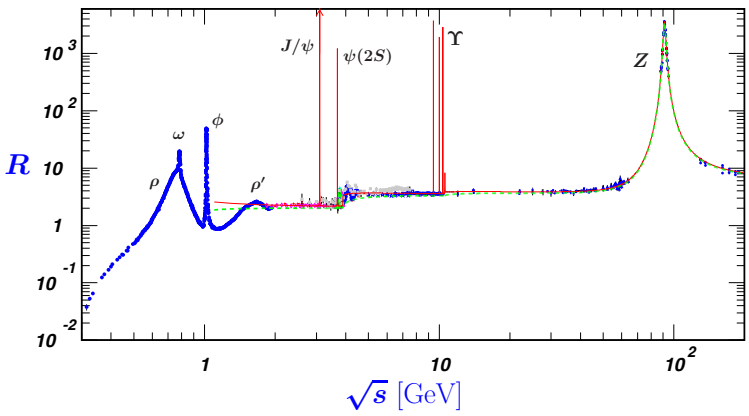

Figure 1.7: World data for the R ratio, of Eq. (1.13), as a function of the center-of-mass energy of the collision system, provided by the PDG [Zyl20]; see [ELZ03] for the original source.

where we note that this is the LO expression. The consequences of this exceedingly-simple computation are highly consequential. The famous measurement of this ratio R, plotted as a function of the center-of-mass energy of the collision, is displayed in Fig. 1.7, where we see that the plateau leading up to the Z^0 resonance is very well-approximated by ~ 3.67 whose value, being directly proportional to N_c, is direct experimental confirmation for QCD having three colors and gauged under $\mathrm{SU}(3)$. One may of course go further with this computation to include higher-order corrections to achieve the pleasing NLO result

$$R_{\mathrm{NLO}} = R_{\mathrm{LO}} \left(1 + \frac{\alpha_s}{\pi} + \mathcal{O}\left(\alpha_s^2\right)\right) , \tag{1.14}$$

which amounts to a small correction indeed.

Now, based on the high level of inclusivity affiliated with the final state of "hadrons," and recalling our previous discussion on jets, one may reasonably relabel the final state as "jets," i.e. $\sigma(e^+e^- \to \mathrm{hadrons}) = \sigma(e^+e^- \to \mathrm{jets})$, semantically. However, again this is only if no further measurements or phase space restrictions are placed on the final state. Let us demonstrate how this cross section is modified if we consider the following simple case: instead of being inclusive over all final state jets, we consider the situation in which we measure a single jet, of radius R and of energy fraction z with respect to the quark initiating the jet, which by conservation of four-momentum is $\sqrt{s}/2$; s being that of Mandelstam. Note the two important ways in which we are breaking the inclusivity of the previous situation: first, we are enforcing a phase space constraints

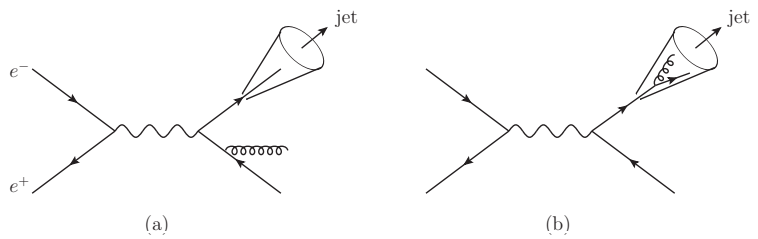

Figure 1.8: Basic kinematic setup for the computation of the quark-initiated semi-inclusive jet function. Figure taken from [KRV16].

by considering our jet to be the collimated beam of hadrons within a conical region of radius R and second, we are performing a measurement of the fraction $z \equiv 2E_J/\sqrt{s}$ with E_J being the energy of the jet.

For values of $z < 1$, the cross section we are after takes the following form [Kre00, ARV13]:

$$\frac{d\sigma}{dz} = N_c\, \sigma_0 \left[F_T^J(z, \sqrt{s}) + F_L^J(z, \sqrt{s}) \right], \tag{1.15}$$

where the functions $F_{T,L}^J$ obey the factorization

$$F_{T,L}^J = \sum_f \left(C_f^{T,L} \otimes J_f \right)(z)\,, \tag{1.16}$$

with $f = q, \bar{q}, g$ and the convolution defined as

$$(g \otimes h)(z) \equiv \int_z^1 \frac{dz'}{z'} g(z') h(z/z')\,. \tag{1.17}$$

The coefficients $C_f^{T,L}$ are capture the physics at their natural hard scale $\mu_h \sim \sqrt{s}$, while the jet functions, J_f, capture that at the collinear scale $\mu_J \sim E_J R$. The T, L superscripts stand for "transverse" and "longitudinal" polarizations, respectively, and thus correspond to the cases where the interaction is mediated through the exchange of a virtual photon or a Z^0 boson, respectively. The precise form of the coefficients $C_f^{T,L}$ are not important here, but their expressions may be found in [Kre00].

The jet functions J_f are known as the semi-inclusive jet functions [DKL16, KRV16] which are constructed through the phase space considerations depicted in Fig. 1.8. The quark-initiated jet

function takes the following form

$$
\begin{aligned}
J_q\left(z,\mu^2\right) = \delta(1-z) &+ \frac{\alpha_s}{2\pi}\left(P_{qg\leftarrow q}(z) + P_{gq\leftarrow q}(z)\right)\log\frac{\mu^2}{E_J^2R^2} \\
&- \frac{\alpha_s}{2\pi}C_F\left\{\left[2(1+z^2)\left(\frac{\log(1-z)}{1-z}\right)_+ + (1-z)\right]\right. \\
&\left. + 2\,\frac{1+(1-z)^2}{z}\log(1-z) + z - \left(\frac{13}{2} - \frac{2\pi^2}{3}\right)\delta(1-z)\right\}, \qquad (1.18)
\end{aligned}
$$

where we have made reference to the Altarelli-Parisi splitting functions

$$
\begin{aligned}
P_{qg\leftarrow q}(z) &= C_F\left(\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\,\delta(1-z)\right), \\
P_{gq\leftarrow q}(z) &= C_F\frac{1+(1-z)^2}{z}\,. \qquad (1.19)
\end{aligned}
$$

$J_q\left(z,\mu^2\right)$ forms the dominant contributor to the process under consideration, see [DKL16, KRV16] for the gluon-initiated case.

The semi-inclusive jet functions satisfy the DGLAP evolution equations

$$
\mu^2\frac{\partial}{\partial\mu^2}\begin{pmatrix} 2n_f J_q\left(z,\mu^2\right) \\ J_g\left(z,\mu^2\right)\end{pmatrix} = \frac{\alpha_s(\mu^2)}{2\pi}\begin{pmatrix} P_{qg\leftarrow q}(z) & 2n_f P_{gq\leftarrow q}(z) \\ P_{q\bar{q}\leftarrow g}(z) & P_{gg\leftarrow g}(z)\end{pmatrix}\otimes\begin{pmatrix} 2n_f J_q\left(z,\mu^2\right) \\ J_g\left(z,\mu^2\right)\end{pmatrix}, \qquad (1.20)
$$

and thus, logarithms of form $\log E_J^2R^2/s$ may be resummed to leading-logarithmic (LL) accuracy by evolving the jet functions up to $\mu \sim s$ through solving Eq. (5.85). Thus, we see that the measurement of a single semi-inclusive jet in the final state of electron-positron collisions is strikingly analogous to the measurement of a parton-distribution function (PDF) in the initial state of deep inelastic scattering (DIS), which is certainly the standard for most introductions to QCD found in textbooks. Fig. 1.9 shows a measurement of the differential cross section $d\sigma/dz$ as factorized in Eq. (1.15), performed with archived ALEPH data [Che22, CLM21]. This concludes our brief introduction to QCD and jets.

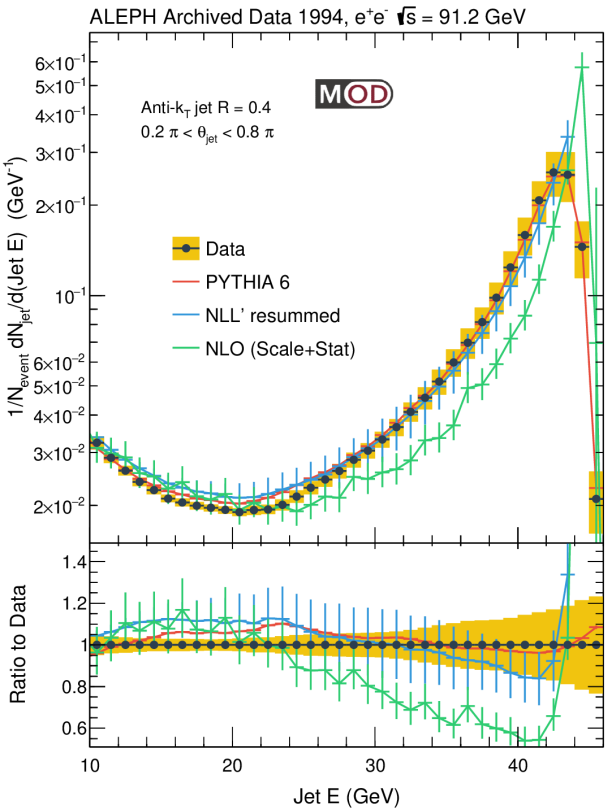

Figure 1.9: Experimental measurement of the energy spectrum of jets of radius $R = 0.4$ produced in electron-positron scattering using archive data taken by the ALEPH Collaboration [Che22, CLM21].

Part I

Heavy quark dijets

CHAPTER 2

Heavy flavor dijet production and dijet mass modification in heavy-ion collisions

---

In this chapter, we utilize back-to-back dijet pairs to probe the quark-gluon plasma created in the final state of heavy-ion collisions. In particular, we analyze the flavor-dependence of the energy-loss mechanisms at play within the dense QCD medium by contrasting the relatives degrees of modification experienced by light and heavy-flavored dijet observables. We define a new observable, the nuclear modification of the dijet invariant mass spectrum, which has the ability to reveal striking differences in suppression patterns between dijets initiated by light partons and those initiated by heavy quarks. We argue that this observable is superior to a traditional observable, known as the dijet momentum imbalance shift, when it comes to the exhibition of such heavy-quark mass-dependence in jet quenching. We present results for Pb+Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV in order to compare to data from the Large Hadron Collider, as well as for Au+Au collisions at $\sqrt{s_{NN}}$ = 200 GeV to guide measurements taken by the sPHENIX experiment at the Relativistic Heavy Ion Collider.

2.1 Introduction

The collision of ultra-relativistic nuclei at facilities such as the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC) result in the formation of a hot and dense medium known as the quark-gluon plasma (QGP). This strongly-interacting state of matter permeates the

final state, and thereby interacts with the jets emerging from the very same collision. As jets evolve and traverse this plasma, they undergo energy loss due to radiative bremsstrahlung as well as elastic collisions with the medium constituents. Such energy loss is referred to as jet quenching. The modification of the production cross sections and substructure of jets is more sensitive to the in-medium strong interaction dynamics in comparison to the leading hadron attenuation [VWZ08a]. As such, jets are excellent diagnostics of the hot and dense medium created in heavy ion collisions (HIC). These jet quenching phenomena have been widely studied at both RHIC and the LHC, for a recent review of jet physics in HIC see Ref. [CNR18].

Heavy flavor jets have been proposed as a new tool to test the theory of heavy quark production, parton shower formation, and modification in nuclear matter. The first theoretical study of single inclusive b-jet production in HIC [HKV13a] has found that the cross section receives a large contribution from hard gluons, where heavy flavor emerges from gluon splitting only in the late stages of the parton shower evolution. Thus, the suppression of single inclusive b-jets at high transverse momenta can be nearly as large as the quenching of light jets, as confirmed by the first CMS measurement [Cha14]. Therefore, the connection between b-jets suppression and b-quark energy loss can be quite indirect. On the other hand, B-meson-tagged b-jets [HKV15a] suppress such a contribution from gluon splitting, and are most effective in ensuring that the dominant fraction of recoiling jets originate from hard b-quarks. Such a conclusion also applies to the back-to-back b-tagged dijet production, as we will show below.

Back-to-back jet pair (or dijet) production is among the most exciting channels used to probe QGP properties, where one typically focuses on the most energetic ("leading") and second most energetic ("subleading") jets. In fact, the first definitive measurement of quenching effects on reconstructed jets came from the enhanced dijet asymmetry measurements at the LHC [Aad10, Cha11]. Further studies of this observable have been carried out not only at the LHC [Kha16], but also at RHIC [Ada17]. The origin of the additional imbalance to the dijet transverse momentum distribution in heavy ion collisions in comparison with the elementary p+p collisions has been attributed to the path length and color charge dependence of parton energy loss [QM11, YSJ11, HVZ12]. The first measurement of the back-to-back b-jet momentum imbalance [Sir18] has been performed at the LHC and modeled theoretically [DWZ18a].

The dijet asymmetry and momentum imbalance measure the *difference* of potentially large attenuation effects on the leading and subleading jets. Thus, those observables show a somewhat reduced sensitivity to the physics of jet quenching and the transport properties of the QGP. To ameliorate this, we set out to find an observable where the effects that arise from the in-medium modification of parton showers *combine* rather than subtract, and thus lead to an enhanced sensitivity to the interactions of jets in the QGP, as well as the mass-dependence of parton energy loss.

In this chapter, we provide an extensive study of dijet production in heavy ion collisions at RHIC (or sPHENIX) kinematics and at LHC energies for both inclusive and b-tagged dijets. We compare the similarities and differences between those channels in A+A collisions to understand the flavor dependence of the quenching effects. Most importantly, we propose to use the dijet invariant mass modification as a novel probe of the QGP. At the same time, we include the studies for the more conventional observables such as two-dimensional nuclear modification factor R_{AA} as a function of leading and subleading jet transverse momenta, and the imbalance z_J distribution. We present theoretical predictions at $\sqrt{s_{NN}}$ = 200 GeV for future Au+Au collisions relevant to the sPHENIX kinematics at RHIC and at $\sqrt{s_{NN}}$ = 5.02 TeV for comparison to Pb+Pb data at the LHC.

This chapter is organized as follows. In Sec. II, we present the evaluation of the differential cross sections for both inclusive and b-tagged dijet production in p+p collisions using the Pythia 8 event generator [SMS08]. We also determine the flavor origin of the dijet production for the proper implementation of the energy loss effects. In Sec. III, we first present the basic formalism used to generate dijet invariant mass distributions and imbalance distributions, starting from the double differential cross section in terms of the transverse momenta of leading and subleading jets. We then provide information regarding our implementation of medium effects to obtain the modification of inclusive and b-tagged dijet production in HIC. In Sec. IV, we present phenomenological results for both RHIC and LHC kinematics. We give predictions for sPHENIX at RHIC, and provide detailed comparison with experimental measurements by the CMS collaboration at the LHC. We conclude in Sec. V.

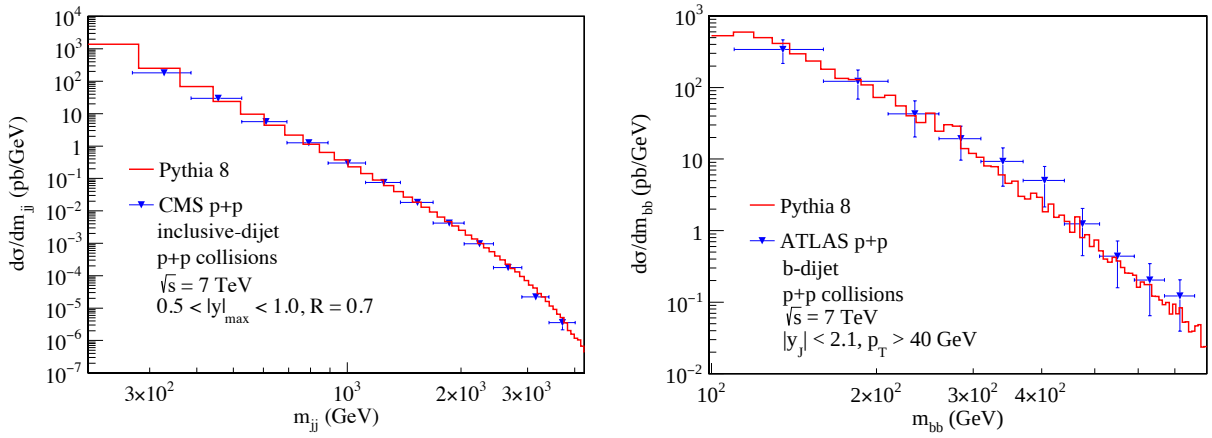

Figure 2.1: Comparison of dijet mass distributions between Pythia 8 simulations and experimental measurements in p+p collisions at the LHC at $\sqrt{s} = 7$ TeV. The left is for inclusive dijets from CMS collaboration [Cha13], while the right is for b-tagged dijets from ATLAS collaboration [Aad11].

2.2 Light and heavy flavor dijet production in p+p collisions

In this section, we present the evaluation of the double differential cross sections for inclusive and b-tagged dijet production in p+p collisions using Pythia 8 [SMS08], which is a widely-used high energy phenomenology event generator that describes the main properties of the event structure well. In our simulations, 8 million events are simulated for each of these two processes. We construct jets with the anti-k_T jet clustering algorithm [CSS08], where a b-jet is identified if there is at least one b-quark within the jet.

Both inclusive and b-tagged dijet production in p+p collisions have been measured at the LHC. Therefore, we begin by validating our Pythia 8 simulations against the experimental measurements of inclusive dijet production in p+p collisions. We present the dijet-production cross section as a function of dijet invariant mass in the left panel of Fig. 2.1, with comparison to experimental measurements by the CMS [Cha13] collaboration at center-of-mass energy $\sqrt{s} = 7$ TeV at the LHC. The dijet invariant mass m_{jj} is simply defined as

$$m_{jj}^2 = (p^{\mathrm{L}} + p^{\mathrm{S}})^2, \tag{2.1}$$

with p^{L} and p^{S} being the four-momenta for the leading and subleading jets, respectively. The jets

are constructed with a jet radius $R = 0.6$, along with the following rapidity cut

$$0.5 < y^* < 1.0, \tag{2.2}$$

where $y^* = |y^{\rm L} - y^{\rm S}|$ with y^L (y^S) being the rapidity of leading and subleading jets. We implement additional cuts on the transverse momentum and rapidity of individual jets, which are matched to those given in the experimental paper [Cha13]. The red histograms are the results from Pythia 8 simulations. As one can see, the Pythia 8 event generator describes the experimental dijet invariant mass data rather well. This allows us to reliably extract information regarding the flavor origins of dijets produced in heavy ion collisions.

In the right panel of Fig. 2.1, we compare our Pythia 8 simulation for b-tagged dijet invariant mass distribution with the ATLAS measurement [Aad11] at $\sqrt{s} = 7$ TeV. The jet radius is $R = 0.4$ and the distance in rapidity and azimuthal angle between b-quark and the b-jet, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, is required to be smaller than 0.3. Additionally, the transverse momentum of the b-quark is required to satisfy $p_T > 5$ GeV. All other event selection and kinematic cuts are implemented to match the experimental measurements. For details, see Ref. [Aad11]. Again, we obtain satisfactory agreement between our Pythia 8 simulation and the experimental data.

With our simulations validated, we next present the detailed baseline information for b-tagged and inclusive dijet production in p+p collisions, at $\sqrt{s} = 5.02$ TeV for the LHC and $\sqrt{s} = 200$ GeV for RHIC. These are the same center-of-mass energies (per nucleon pair) for the current heavy ion collisions at the LHC and for the sPHENIX experiment, respectively.

In Fig. 2.2, we show the three-dimensional (3D) plots of the cross section (weighted by the transverse momenta $p_{1T}p_{2T}$) at the LHC energy $\sqrt{s} = 5.02$ TeV as a function of the transverse momenta of the two jets (p_{1T} and p_{2T}) in the mid-rapidity region $|y| < 2$. The jets are reconstructed with a jet radius $R = 0.4$. Here we label the dijet transverse momenta as p_{1T} and p_{2T} (instead of $p_T^{\rm L}$ and $p_T^{\rm S}$), because we do not distinguish which jet is leading or subleading in the generation of the 3D plots. We will follow this same convention in the rest of the chapter—when we need to specify leading and subleading jets, we label them as $p_T^{\rm L}$ and $p_T^{\rm S}$. Otherwise, we simply label them as p_{1T} and p_{2T}. The left plot is for b-tagged dijet production, while the right is for inclusive dijets.

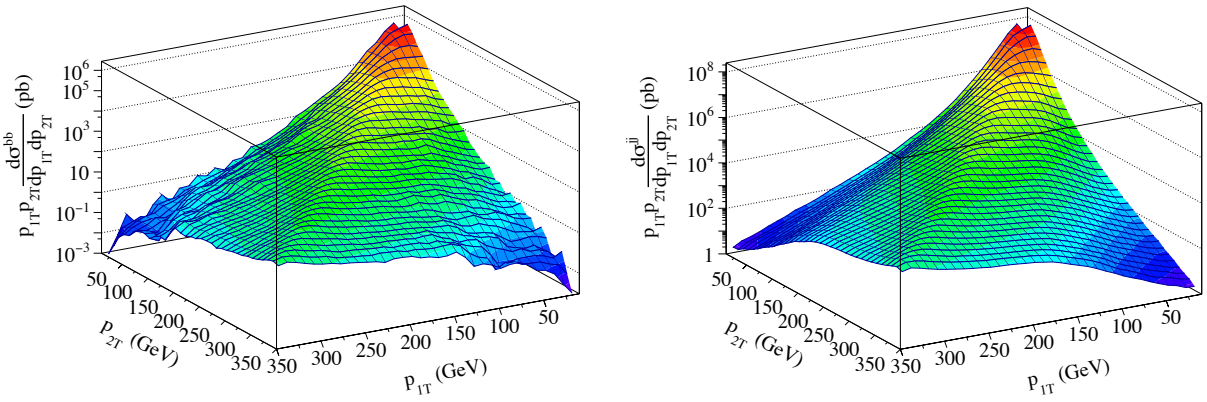

Figure 2.2: Double differential cross sections weighted by transverse momenta for b-tagged (left) and inclusive (right) dijet production in p+p collisions at $\sqrt{s} = 5.02$ TeV. Kinematic cuts are implemented in our simulations as in CMS measurements, see Ref. [Sir18]. The roughness of the b-tagged dijet cross section relative to that for inclusive dijets is due to the inherently lower statistics.

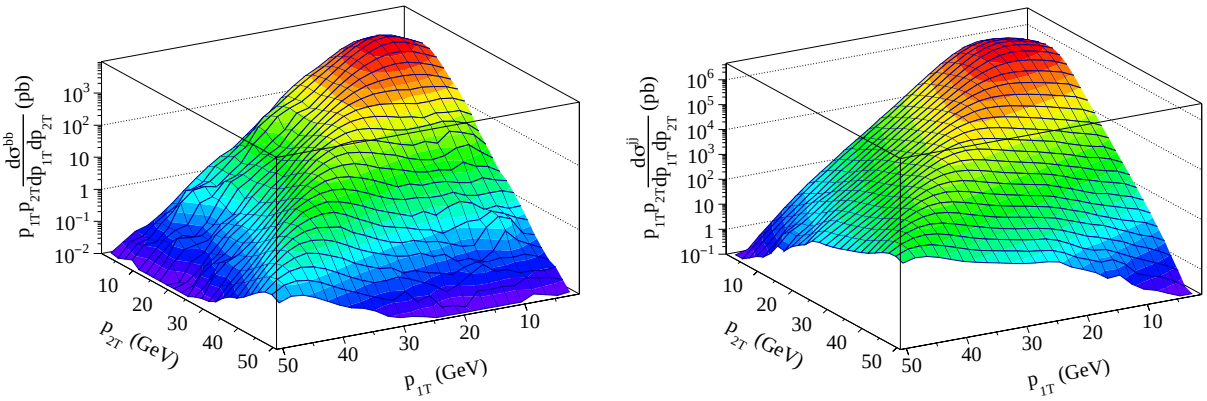

Figure 2.3: Double differential cross sections weighted by transverse momenta for b-tagged (left) and inclusive (right) dijet production in p+p collisions at $\sqrt{s} = 200$ GeV. Kinematic cuts implemented in our simulations are the same as those from the sPHENIX collaboration [sPH]. Here again, the slight bumpiness of the b-tagged dijet cross section is due to its lower statistics relative to the inclusive case.

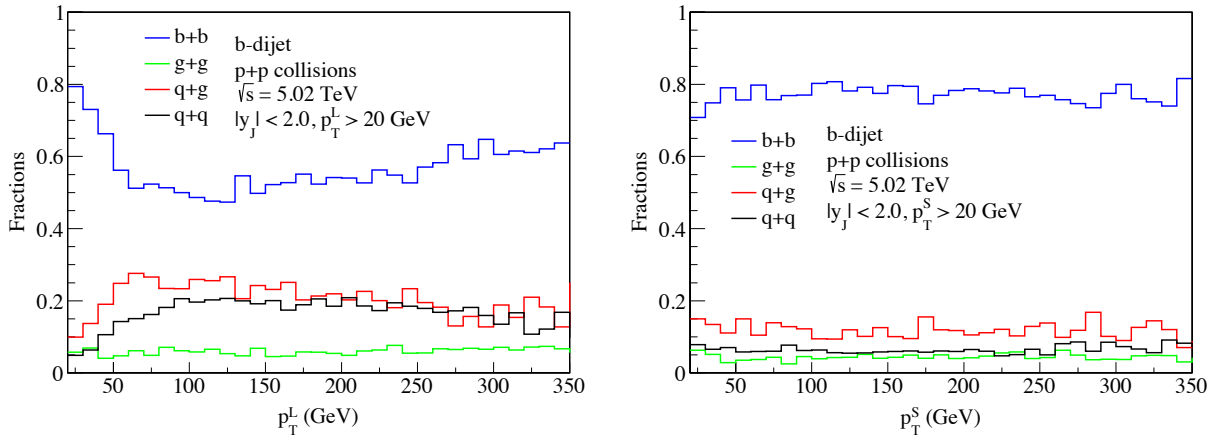

Figure 2.4: The fractional contributions of different subprocesses to the b-dijet production cross sections vs. leading jet $p_T^{\rm L}$ (left) and subleading jet $p_T^{\rm S}$ (right) in p+p collisions at $\sqrt{s} = 5.02$ TeV. Kinematic cuts are implemented in our simulations as in CMS measurements [Sir18].

Fig. 2.3 is the same as Fig. 2.2, but for sPHENIX energy $\sqrt{s} = 200$ GeV. The roughness of the b-tagged dijet cross section relative to that for inclusive dijets is due to the inherently lower statistics. As usual [HVZ12], the cross section reaches its maximum for $p_{1T} \approx p_{2T}$, and is broad and slowly varying outside of the main diagonal. Such features will help us understand the behavior of nuclear modification in heavy ion collisions as we will see below.

Let us now turn to the flavor origin of the dijets, which will not only be of central importance for our simulations of medium effects in heavy ion collisions (presented in the next section), but will also underpin the efficacy of b-dijets as probes of b-quark energy loss. The detailed kinematic constraints are shown in each plot. Pythia 8 utilizes leading order (LO) perturbative QCD matrix elements combined with parton showers. For b-tagged dijet production, there are 7 channels in our simulations: $g+g \to b+\bar{b},\ q+\bar{q} \to b+\bar{b},\ g+g \to g+g,\ q+\bar{q} \to g+g,\ q+g \to q+g,\ g+g \to q+\bar{q}, q+q \to q+q$. We classify these 7 channels to 4 subprocesses according to the flavor information of the final state partons in LO matrix elements: (1) $g+g \to b+\bar{b}$, $q+\bar{q} \to b+\bar{b}$; (2) $g+g \to g+g$, $q+\bar{q} \to g+g$; (3) $q+g \to q+g$; (4) $g+g \to q+\bar{q}$, $q+q \to q+q$. We show in Figs. 2.4 and 2.5 the fractions of these 4 subprocesses as functions of leading (trigger) jet $p_T^{\rm L}$ and subleading (associate) jet $p_T^{\rm S}$ at $\sqrt{s} = 5.02$ TeV and $\sqrt{s} = 200$ GeV, respectively.

The blue line labeled as $b + \bar{b}$ denotes the contributions from category (1), with $b + \bar{b}$ in the

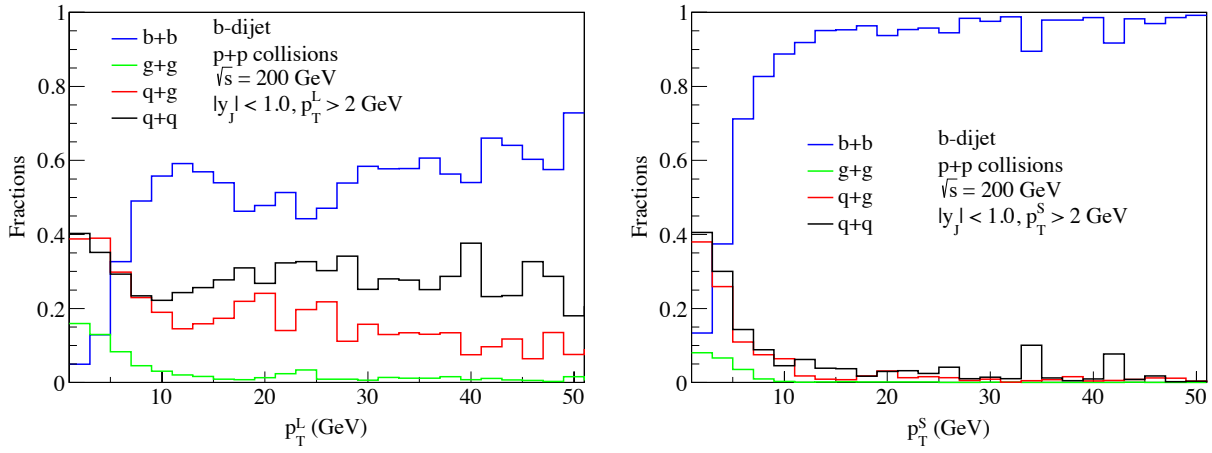

Figure 2.5: The fractional contributions of different subprocesses to the b-dijet production cross sections vs. leading p_T (left) and subleading p_T (right) in p+p collisions at $\sqrt{s} = 200$ GeV. Kinematic cuts implemented in our simulations are the same as those from the sPHENIX collaboration [sPH].

final state of the hard collision. In this case, both b-tagged jets are initiated by either a b-quark or a $\bar{b}$-quark. In heavy ion collisions, the medium modification of such b-jets has a direct connection to the physical heavy quark energy loss (mass m_b). The green curve labeled as $g + g$ includes the contributions from category (2), with $g + g$ in the final state. In this case, both b-tagged jets are initiated by prompt gluons through $g \to b + \bar{b}$ splitting in the showering process. Thus, the medium modification of these b jets would resemble that of a massive gluon of effective mass $2m_b$. Similarly, the red curve denotes the process from category (3). Thus, one b-jet is initiated by a gluon g like above. On the other hand, the other b-jet is initiated by a light quark q, for which the medium modification would resemble that of a massive quark. Finally, the black curve denotes the processes in category (4). In this case, both of the b-tagged jets are initiated by light quarks q. As we can see, for a wide kinematic coverage, the subprocesses with $b + \bar{b}$ in the final state provide the dominant contributions ($\gtrsim 50\%$) to b-tagged dijet production in p+p collisions at the LHC at $\sqrt{s} = 5.02$ TeV. On the other hand, the $b+\bar{b}$ channel dominates b-tagged dijet production across the p_T range above 10 GeV, which is the relevant range for the sPHENIX experiment. This indicates that b-tagged dijet production provides an excellent opportunity to study the effects of heavy quark energy loss in heavy ion collisions.

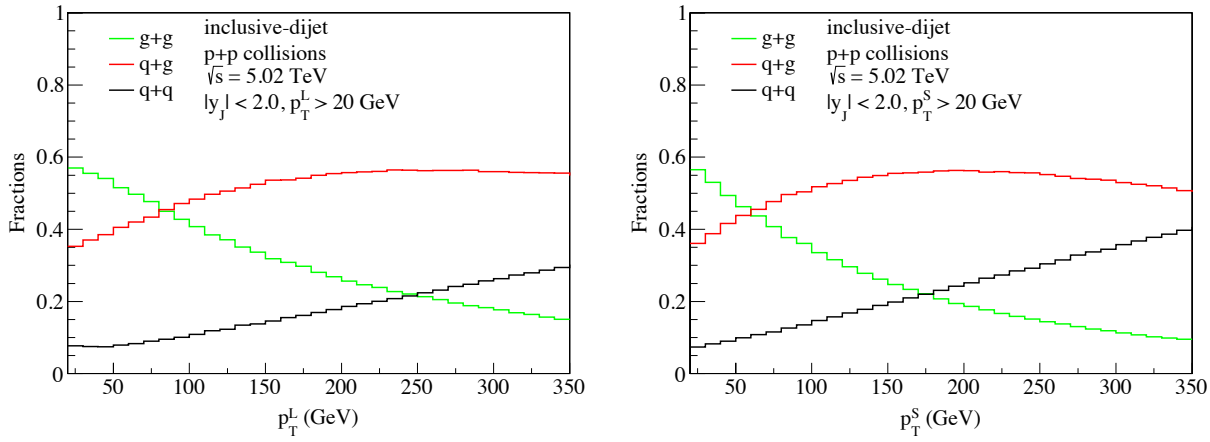

Figure 2.6: The fractional contributions of different subprocesses to the inclusive dijet production cross sections vs. leading p_T (left) and subleading p_T (right) in p+p collisions at $\sqrt{s} = 5.02$ TeV. Kinematic cuts are implemented in our simulations as in CMS measurements [Sir18].

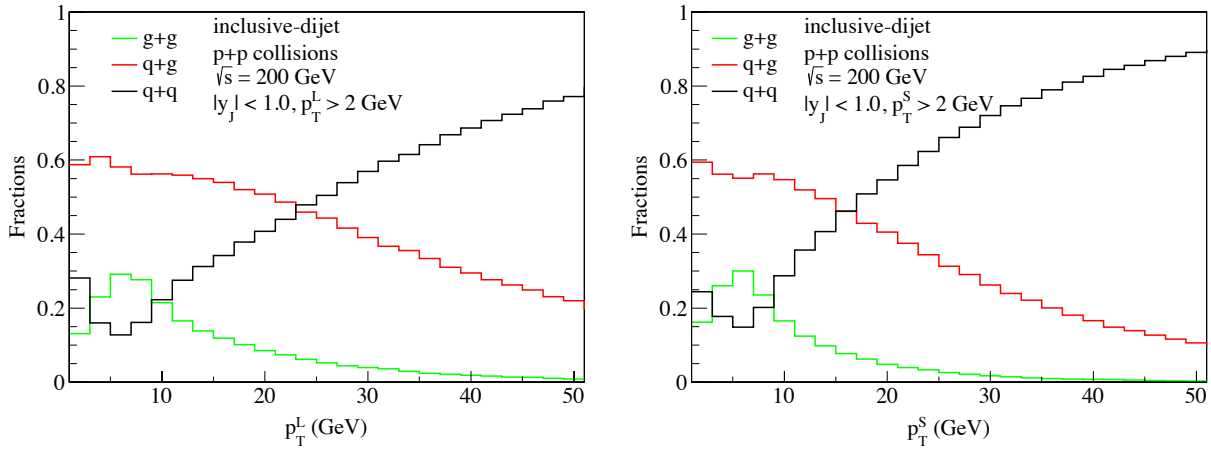

Figure 2.7: The fractional contributions of different subprocesses to the inclusive dijet production cross sections vs. leading p_T (left) and subleading p_T (right) in p+p collisions at $\sqrt{s} = 200$ GeV. Kinematic cuts implemented in our simulations are the same as those from the sPHENIX collaboration [sPH].

On the other hand, for inclusive dijet production, the usual 5 partonic processes will be reclassified into three subprocesses through their final state parton contents: (1) $g+g \to g+g,\ q+\bar{q} \to g+g$; (2) $q+g \to q+g$; (3) $g+g \to q+\bar{q}, q+q \to q+q$. In category (1), both jets are initiated by gluons, while for category (3), both jets are initiated by quarks. For category (2), the dijets are initiated by a light quark q and a gluon g, respectively. One can clearly see in Fig. 2.6 that at LHC energy $\sqrt{s} = 5.02$ TeV, for a large kinematic region, the process from category (2) is the dominant channel for inclusive dijet production. In other words, inclusive dijets at LHC kinematics are mostly initiated by a quark q on one side and a gluon g on the other end of the azimuthal plane. In addition, we plot such fractions in Fig. 2.7 at sPHENIX energy $\sqrt{s} = 200$ GeV as a function of leading jet transverse momentum $p_T^{\rm L}$ (left panel) and of the subleading jet transverse momentum $p_T^{\rm S}$ (right panel), respectively. We find that at relatively lower jet transverse momenta ($\lesssim 20$ GeV), the inclusive dijet cross section is dominated by category (2) with $q+g$ in the final state. At the higher jet transverse momenta, the cross section is dominated by category (3) with $q+q$ in the final state. This is expected since as the jet transverse momenta increase, the parton momentum fractions x in the protons reach the region $x \sim 1$, where valence quarks dominate.

2.3 Light and heavy flavor dijet production in hot QCD matter

In this section, we provide the main formula and basic information on how we implement parton energy loss for both inclusive and b-tagged dijet production in heavy ion collisions.

2.3.1 Dijet production: main formula

Our starting point for both p+p and A+A collisions is the double differential cross section, $d\sigma/dp_{1T}dp_{2T}$, in two-dimensional transverse momentum bins (p_{1T}, p_{2T}) of the leading and subleading jets. With such a double differential cross section at hand, one can compute the dijet invariant mass distribution, as well as the so-called imbalance distribution as follows.

The dijet invariant mass $m_{12}^2 = (p_1+p_2)^2$ can be written in terms of the jet transverse momen-

tum and rapidity as follows

$$m_{12}^2 = m_1^2 + m_2^2 + 2\left[m_{1T}m_{2T}\cosh(\Delta\eta) - p_{1T}p_{2T}\cos(\Delta\phi)\right], \tag{2.3}$$

where $m_1^2 = p_1^2$ and $m_{1T} = \sqrt{m_1^2 + p_{1T}^2}$ are the invariant mass squared and the transverse mass for one of the jets, likewise we have m_2 and m_{2T} for the other jet. At the same time, we have the difference in the rapidities and the azimuthal angles as

$$\Delta\eta = \eta_1 - \eta_2, \qquad \Delta\phi = \phi_1 - \phi_2, \tag{2.4}$$

where $\eta_{1,2}$ and $\phi_{1,2}$ are the rapidities and azimuthal angles for the jets. In the relevant kinematic regimes where the transverse momentum is much larger than the jet mass, $p_T \gg m$, we approximate $m_T \approx p_T$ and obtain

$$m_{12}^2 \approx m_1^2 + m_2^2 + 2p_{1T}p_{2T}\left[\cosh(\Delta\eta) - \cos(\Delta\phi)\right]. \tag{2.5}$$

In the actual Pythia 8 simulations for dijet production, we generate the averaged $\langle m_1^2\rangle$, $\langle m_2^2\rangle$, and $\langle\cosh(\Delta\eta) - \cos(\Delta\phi)\rangle$ for each (p_{1T}, p_{2T}) bin. With this information, we compute the dijet invariant mass distribution through the double differential dijet momentum distribution via the following formula

$$\frac{d\sigma}{dm_{12}} = \int dp_{1T}dp_{2T}\frac{d\sigma}{dp_{1T}dp_{2T}}\delta\left(m_{12} - \sqrt{\langle m_1^2\rangle + \langle m_2^2\rangle + 2p_{1T}p_{2T}\langle\cosh(\Delta\eta) - \cos(\Delta\phi)\rangle}\right), \tag{2.6}$$

where the transverse momenta p_{1T} and p_{2T} are integrated over the desired experimental cuts.

Let us now confirm that such a procedure yields correct dijet mass distributions. To do this, we compare the dijet invariant mass distribution *indirectly* computed using Eq. (2.6) and Pythia 8-simulated $d\sigma/dp_{1T}dp_{2T}$, with $d\sigma/dm_{12}$ simulated *directly* from Pythia 8. We perform such a comparison in Figs. 2.8 and 2.9 for b-tagged (left panels), as well as for inclusive (right panels), dijet production at LHC energy $\sqrt{s} = 5.02$ TeV and sPHENIX energy $\sqrt{s} = 200$ GeV, respectively. In the top panels, the black histograms represent $d\sigma/dm_{12}$ simulated *directly* from Pythia 8, while the red histograms are $d\sigma/dm_{12}$ computed using Eq. (2.6) and Pythia 8-simulated $d\sigma/dp_{1T}dp_{2T}$. In the bottom panels, the black histograms mark the baseline at unity for the mass distribution ratios

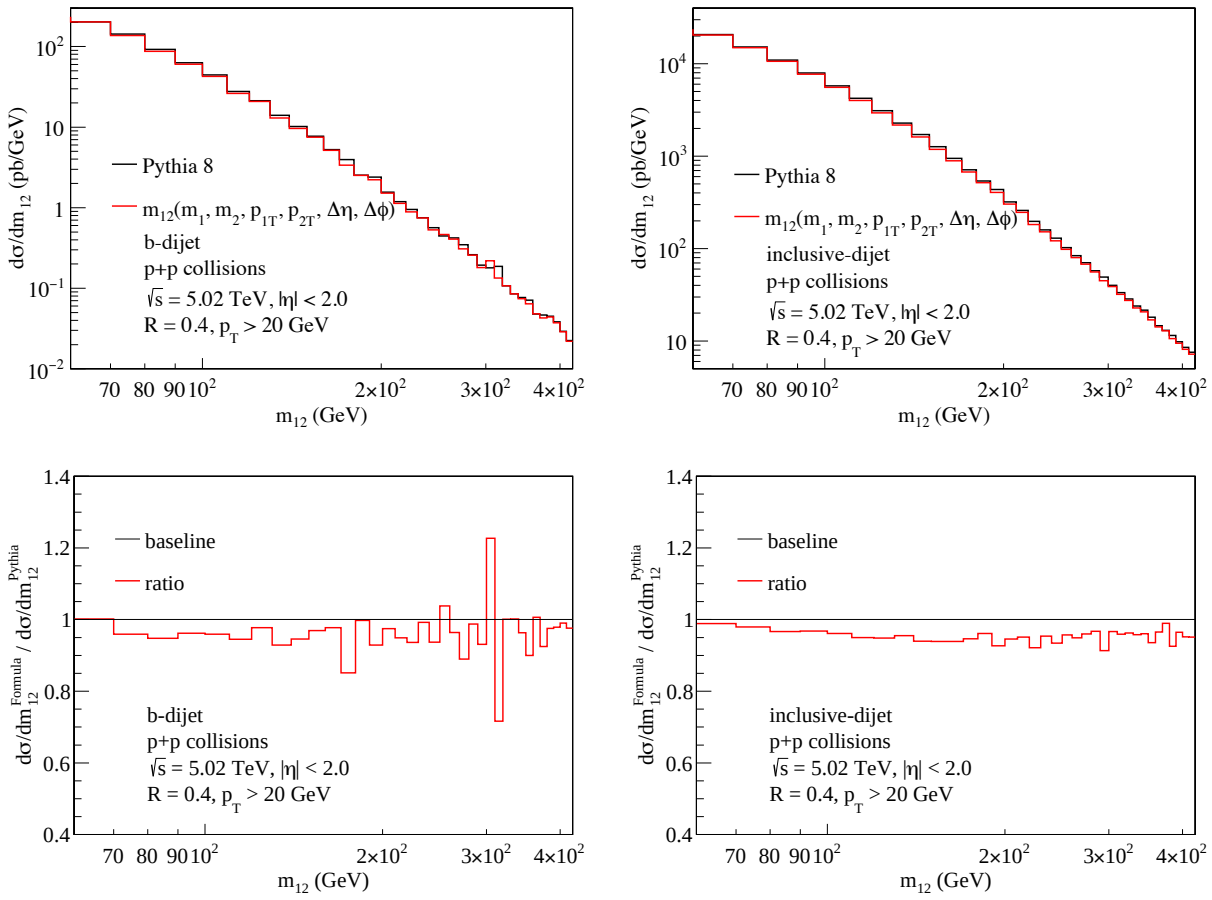

Figure 2.8: Mass distributions (top) and their ratios (bottom) for b-tagged (left) and inclusive (right) dijet production in p+p collisions at $\sqrt{s} = 5.02$ TeV. Kinematic cuts are implemented in our simulations as in CMS measurements [Sir18]. The upper panels display black histograms representing $d\sigma/dm_{12}$ simulated *directly* from Pythia 8, while the red histograms are $d\sigma/dm_{12}$ computed using Eq. (2.6) and Pythia 8-simulated $d\sigma/dp_{1T}dp_{2T}$. The lower panels display black histograms denoting the baseline ratio for the mass distributions (representing good agreement) with red histograms representing the actual simulated ratio.

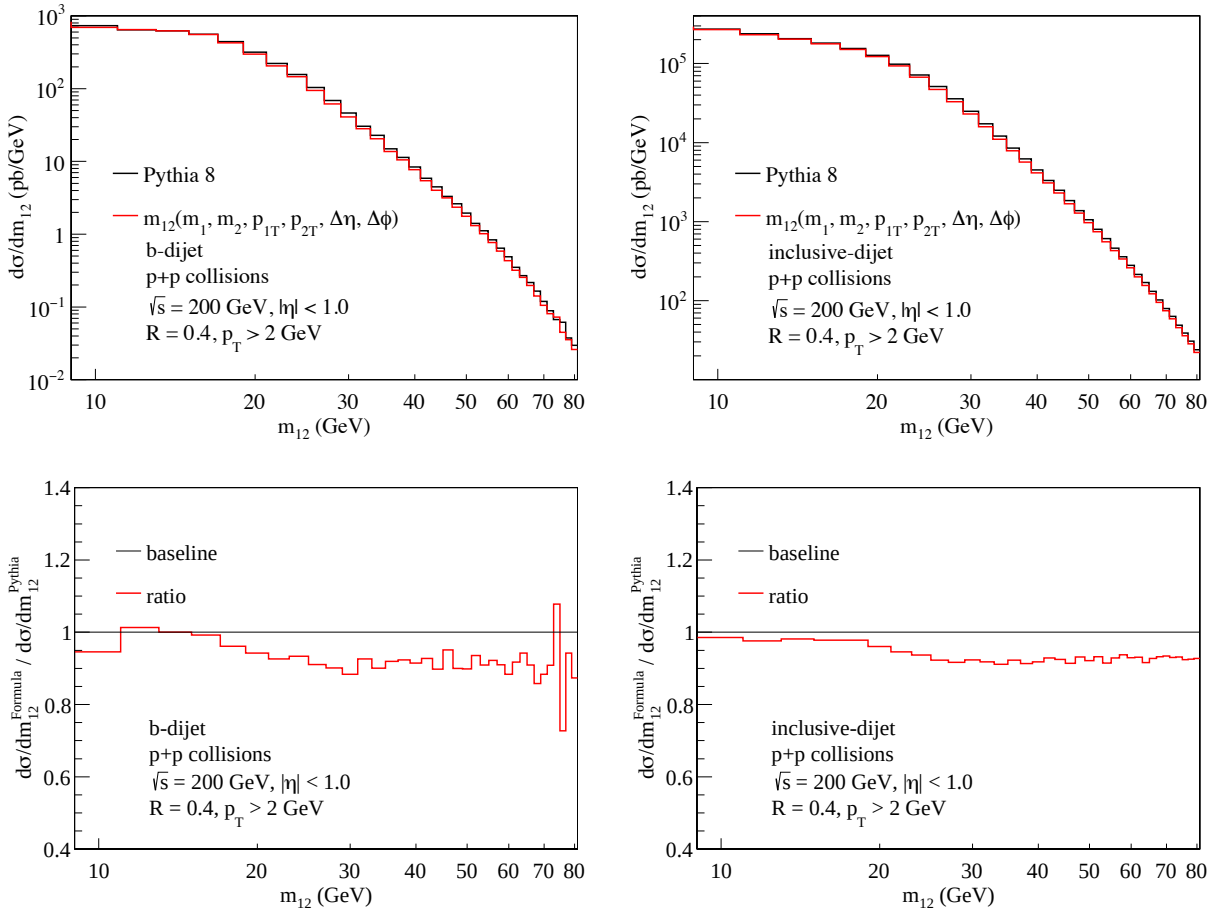

Figure 2.9: Mass distributions (top) and their ratios (bottom) for b-tagged (left) and inclusive (right) dijet production in p+p collisions at $\sqrt{s} = 200$ GeV. Kinematic cuts implemented in our simulations are the same as those from the sPHENIX collaboration [sPH]. The upper panels display black histograms representing $d\sigma/dm_{12}$ simulated *directly* from Pythia 8, while the red histograms are $d\sigma/dm_{12}$ computed using Eq. (2.6) and Pythia 8-simulated $d\sigma/dp_{1T}dp_{2T}$. The lower panels display black histograms denoting the baseline ratio for the mass distributions (representing good agreement) with red histograms representing the actual simulated ratio.

while the red histograms represent the ratio between the mass distributions utilizing Eq. (2.6) and those directly from Pythia 8. We observe a quite reasonable matching of mass spectra obtained via direct implementation of dijet mass in Pythia 8 and our approximate formula in Eq. (2.6), as indicated by the fact that the approximate distributions only represent a $\sim 10\%$ undercounting. This induces a minor change in the overall normalization of each distribution whose effect cancels out in the computation of the nuclear modification factor R_{AA}. This validates the use of our formula in applications to heavy-ion collisions and subsequent dijet mass modifications.

On the other hand, one of the more conventional observables, the dijet momentum imbalance shift, is based on the cross section as a function of the imbalance variable

$$z_J = p_{2T}/p_{1T}, \tag{2.7}$$

which can be derived from the double differential cross section $d\sigma/dp_{1T}dp_{2T}$. The formula is given as follows

$$\frac{d\sigma}{dz_J} = \int dp_{1T}dp_{2T}\frac{d\sigma}{dp_{1T}dp_{2T}}\delta\left(z_J - \frac{p_{2T}}{p_{1T}}\right), \tag{2.8}$$

where again, the limits of integration for p_{1T} and p_{2T} are matched to the desired experimental cuts.

Comparing Eqs. (2.6) with (2.8), one can gain some insights why medium modification of dijet invariant mass distribution leads to enhanced medium effects than that of the dijet momentum imbalance. This is because dijet invariant mass $m_{12} \propto p_{1T}p_{2T}$, i.e. product of two jet momenta, and thus leads to a combination of the jet quenching effects on the individual jets. On the other hand, the momentum imbalance $z_J = p_{2T}/p_{1T}$, i.e. quotient of two jet momenta, and thus leads to an subtraction of the jet quenching effects on the individual jets. We elaborate more on this point in the presentation of our numerical results below.

2.3.2 Modification of dijet production

In the presence of the hot and dense QCD medium, the vacuum parton shower gets modified due to the radiative [Zak97, BDM97, GLV00, Wie00, WG01, AMY02, ZWW04, DK01] and collisional [BT91, WHD07, AGH07, Tho09, BBC14, NVX14] energy losses of the propagating partons that initiate and form the jets. The implementation of energy loss effects in heavy ion collisions is

explained in detail in, e.g., Refs. [HVZ12, KVX17]. We will give a general overview of the setup in what follows. For a given impact parameter $|\boldsymbol{b}_\perp|$ in the transverse plane of the nucleus-nucleus collisions, the inclusive dijet double differential cross section in (p_{1T}, p_{2T}) can be expressed as

$$\begin{aligned}\frac{d\sigma^{AA}(|\boldsymbol{b}_\perp|)}{dp_{1T}dp_{2T}} &= \int d^2\boldsymbol{s}_\perp T_A\left(\boldsymbol{s}_\perp - \frac{\boldsymbol{b}_\perp}{2}\right) T_A\left(\boldsymbol{s}_\perp + \frac{\boldsymbol{b}_\perp}{2}\right) \\ &\times \sum_{q,g}\int_0^1 d\epsilon \frac{P^1_{q,g}(\epsilon;\boldsymbol{s}_\perp,|\boldsymbol{b}_\perp|)}{1-f^{1\,\mathrm{loss}}_{q,g}(R;s_\perp,|\boldsymbol{b}_\perp|)\,\epsilon}\int_0^1 d\epsilon' \frac{P^2_{q,g}(\epsilon';\boldsymbol{s}_\perp,|\boldsymbol{b}_\perp|)}{1-f^{2\,\mathrm{loss}}_{q,g}(R;s_\perp,|\boldsymbol{b}_\perp|)\,\epsilon'} \\ &\times \frac{d\sigma^{NN}_{q,g}\left(p_{1T}/[1-f^{1\,\mathrm{loss}}_{q,g}(R;s_\perp,|\boldsymbol{b}_\perp|)\,\epsilon], p_{2T}/[1-f^{2\,\mathrm{loss}}_{q,g}(R;s_\perp,|\boldsymbol{b}_\perp|)\,\epsilon']\right)}{dp_{1T}dp_{2T}}\,, \end{aligned} \tag{2.9}$$

where $|\boldsymbol{b}_\perp|$ is the mean impact parameter for a given collision centrality. In the b-tagged dijet case, we further include the b-quark contributions that initiates prompt b-jets. In Eq. (2.9), $T_A(\boldsymbol{s}_\perp) = \int_{-\infty}^{\infty}\rho_A(\boldsymbol{s}_\perp, z)dz$ is the so-called thickness function in the usual optical Glauber model, where we choose the inelastic nucleon-nucleon scattering cross section $\sigma_{\mathrm{in}} = 70$ mb (42 mb) to obtain average number of binary collisions at $\sqrt{s_{NN}} = 5.02$ TeV (200 GeV) [MRS07], respectively. $P_{q,g}(\epsilon)$ is the probability density for the parent parton to redistribute a fraction ϵ of its energy through medium-induced soft gluon bremsstrahlung. For reconstructed jets, what matters is the out-of-cone energy loss fraction $f^{loss}_{q,g}$ [KVX17]

$$f^{\mathrm{loss}}_{q,g}(R;\mathrm{rad}+\mathrm{coll}) = 1 - \left(\int_0^R dr\int_{\omega_{\min}}^E d\omega\,\frac{dN^g_{q,g}(\omega,r)}{d\omega dr}\right)\Bigg/\left(\int_0^{R_{\max}} dr\int_0^E d\omega\,\frac{dN^g_{q,g}(\omega,r)}{d\omega dr}\right)\,, \tag{2.10}$$

which includes both radiative and collisional energy loss effects, with $\omega_{\min}$ being a parameter that controls the energy dissipated by the medium-induced parton shower into the QGP due to collisional processes [NVX14]. On the other hand, $\frac{dN^g_{q,g}(\omega,r)}{d\omega dr}$ is the medium-induced gluon distribution [Vit07], which is the soft emission limit of the complete in-medium splitting functions [OV12]. For further details regarding the computation and implementation of the above formula, see the original paper [KRV19a].

Once we obtain the medium-modified differential cross section $d\sigma^{AA}/dp_{1T}dp_{2T}$, we then use Eqs. (2.6) and (2.8) to compute the dijet invariant mass distribution $d\sigma^{AA}/dm_{12}$ and imbalance distribution $d\sigma^{AA}/dz_J$ in heavy ion collisions. Such a procedure is perfectly fine for $d\sigma^{AA}/dz_J$, but is an approximation for $d\sigma^{AA}/dm_{12}$, where we assume that the medium modification for the

single jet mass distributions $\langle m_1^2 \rangle$ and $\langle m_2^2 \rangle$ are much smaller than those for the transverse momenta p_{1T} and p_{2T}. Thus, starting from Eq. (2.6), we obtain

$$\frac{d\sigma^{AA}}{dm_{12}} = \int dp_{1T} dp_{2T} \frac{d\sigma^{AA}}{dp_{1T} dp_{2T}} \delta \left(m_{12} - \sqrt{\langle m_1^2 \rangle_{pp} + \langle m_2^2 \rangle_{pp} + 2p_{1T} p_{2T} \langle \cosh(\Delta\eta) - \cos(\Delta\phi) \rangle_{pp}} \right), \tag{2.11}$$

where we have used the same values for $\langle m_1^2 \rangle$, $\langle m_2^2 \rangle$, and $\langle \cosh(\Delta\eta) - \cos(\Delta\phi) \rangle$ as those in p+p collisions, as denoted by the subscript pp. Such an approximation is well-justified. For example, mass distributions for single inclusive jets are indeed not significantly modified, as observed by ALICE collaboration at the LHC [Ach18].

2.4 Phenomenological results at RHIC and the LHC

In this section we first present our phenomenological results for both inclusive and b-tagged dijet production in A+A collisions at the LHC, as well as the future sPHENIX experiment at RHIC. To investigate dijet production in heavy ion collisions and quantify its deviation from the baseline results in elementary p+p reactions, we start with the two-dimensional nuclear modification factor

$$R_{AA}(p_{1T}, p_{2T}, |\boldsymbol{b}_\perp|) = \frac{1}{\langle N_{\rm bin} \rangle} \frac{d\sigma^{AA}(|\boldsymbol{b}_\perp|)/dp_{1T} dp_{2T}}{d\sigma^{pp}/dp_{1T} dp_{2T}}, \tag{2.12}$$

where $|\boldsymbol{b}_\perp|$ is the corresponding impact parameter and $\langle N_{\rm bin} \rangle$ is the average number of nucleon-nucleon scatterings for a given centrality class. The R_{AA} observable can thus be interpreted as a likelihood ratio of probability distributions differential in the transverse momenta of the jets forming the pair. The factor of $1/\langle N_{\rm bin} \rangle$ essentially scales the pp distribution to give it approximately the same normalization as that of the AA.

In Fig. 2.10, we make 3D plots for nuclear modification factor R_{AA} as a function of the jet transverse momenta p_{1T} and p_{2T} simultaneously. The calculations are done for the production of dijets with radii $R = 0.4$ in central $(0 - 10\%)$ Pb+Pb collisions at the LHC energy $\sqrt{s_{NN}} = 5.02$ TeV. We integrate the rapidities of both jets over the interval $|y| < 2$. For the medium effects, we choose the coupling between the jet and the medium to be $g = 1.8$. This is consistent with the value used in our previous studies for single inclusive jets [KRV17b], vector-boson-tagged jets [KVX17],

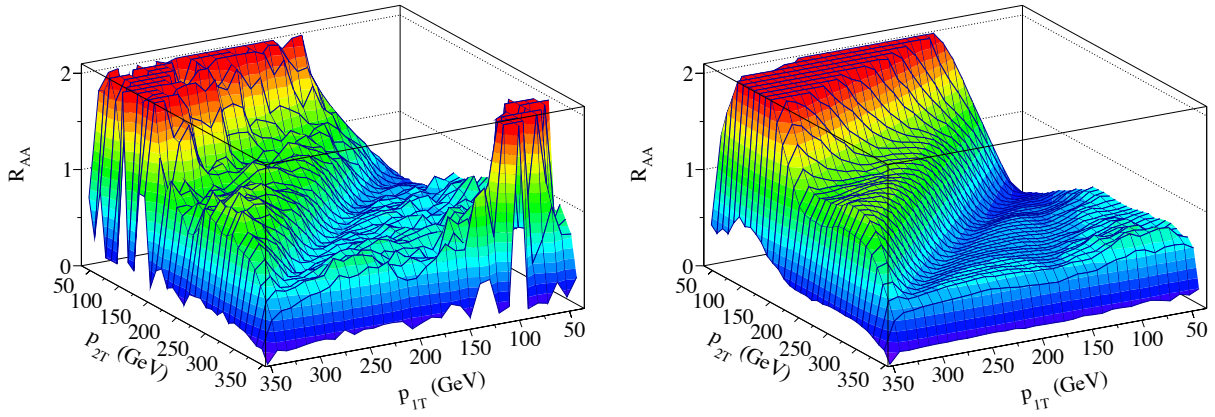

Figure 2.10: Nuclear modification factor for b-tagged (left) and inclusive (right) dijet production in p+p collisions at $\sqrt{s} = 5.02$ TeV. Kinematic cuts are implemented in our simulations as in CMS measurements [Sir18].

jet substructure [CV16a, CV17], and single inclusive hadrons [KLO15, CEK16, KRV17a] in A+A collisions. The left figure is for b-tagged dijet production, while the right is for inclusive dijets. We note that while we plot the full symmetric range in p_{1T} and p_{2T}, we do have in mind that the first jet (1) will be the trigger or leading jet and the second jet (2) will be the recoil or subleading jet. Thus, we incorporate on average path length and color charge bias effects in our calculation.

As one can clearly see, the largest suppression occurs along the diagonal $p_{1T} = p_{2T}$, consistent with our expectation. In the region away from the diagonal, there is a striking enhancement. As the future sPHENIX [Ada15] experiment will have good sensitivity in measuring both inclusive and b-tagged dijet production, it is an opportune time to make predictions for sPHENIX kinematics. In Fig. 2.11 we make similar 3D plots of R_{AA} for b-tagged (left) and inclusive (right) dijet production at sPHENIX energy $\sqrt{s_{NN}} = 200$ GeV. Kinematic cuts implemented in our simulations are the same as those from the sPHENIX collaboration [sPH]. Obviously the kinematic coverage for the jet transverse momenta is much smaller than that of the jets at the LHC, due to a much smaller center-of-mass energy. However, the suppression is even stronger along the diagonal $p_{1T} = p_{2T}$. This is simply because the cross sections at RHIC energies fall much faster as functions of jet transverse momenta due to limited phase space, and thus jet quenching effects get amplified [Vit05, AG04, WW05, Abe07, Ada08].

If such two-dimensional nuclear modification ratios could be measured in detail, they would

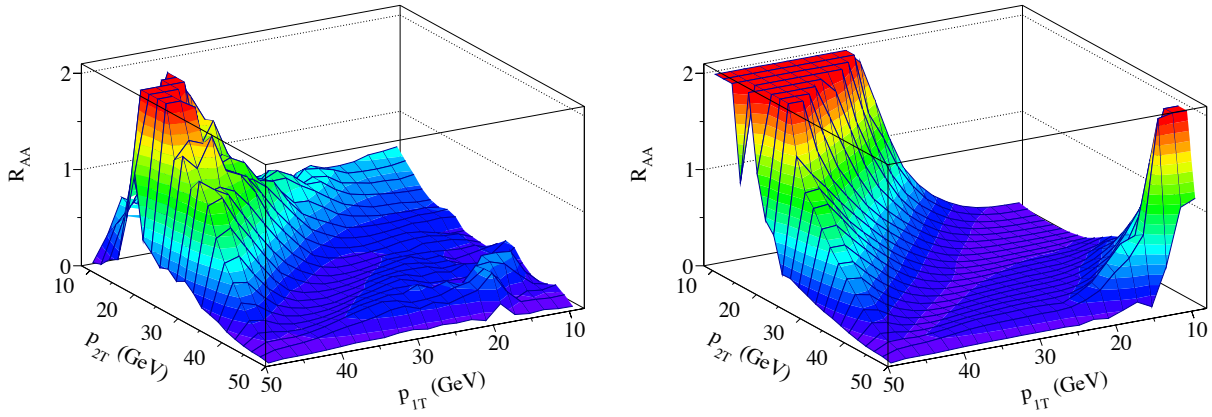

Figure 2.11: Nuclear modification factor for b-tagged (left) and inclusive (right) dijet production in p+p collisions at $\sqrt{s} = 200$ GeV. Kinematic cuts implemented in our simulations are the same as those from the sPHENIX collaboration [sPH].

provide the most information and insight into jet quenching and heavy flavor dynamics in the medium. However, the statistics necessary to perform such measurements make this, at present, quite difficult. In practice, one usually integrates out one of the differential variables and, thus, achieves a one-dimensional nuclear modification ratio. In this respect, the conventional dijet momentum imbalance z_J and asymmetry A_J distributions have been extensively studied in the literature. The medium modification on these traditional distributions emphasize the difference in the quenching of the dijet production, which has been observed to be relatively small. We will present such studies toward the end of this section.

Here instead, we present the nuclear modification for another observable, the dijet invariant mass distribution, defined as follows

$$R_{AA}(m_{12}, |\boldsymbol{b}_\perp|) = \frac{1}{\langle N_{\rm bin}\rangle} \frac{d\sigma^{AA}(|\boldsymbol{b}_\perp|)/dm_{12}}{d\sigma^{pp}/dm_{12}}. \tag{2.13}$$

Again, the impact parameter $|\boldsymbol{b}_\perp|$ indicates the centrality class for the A+A collisions. The numerator and denominator are the dijet mass distribution in A+A and p+p collisions, respectively. They are computed through the double differential cross sections $d\sigma/d_{1T}dp_{2T}$ as in Eqs. (2.11) and (2.6), respectively. In Eqs. (2.6) and (2.11), one can immediately see the advantage of such an observable. First, being only differential in the dijet invariant mass m_{12}, it is a one-dimensional observable, hence one should have enough statistics to perform these measurements experimen-

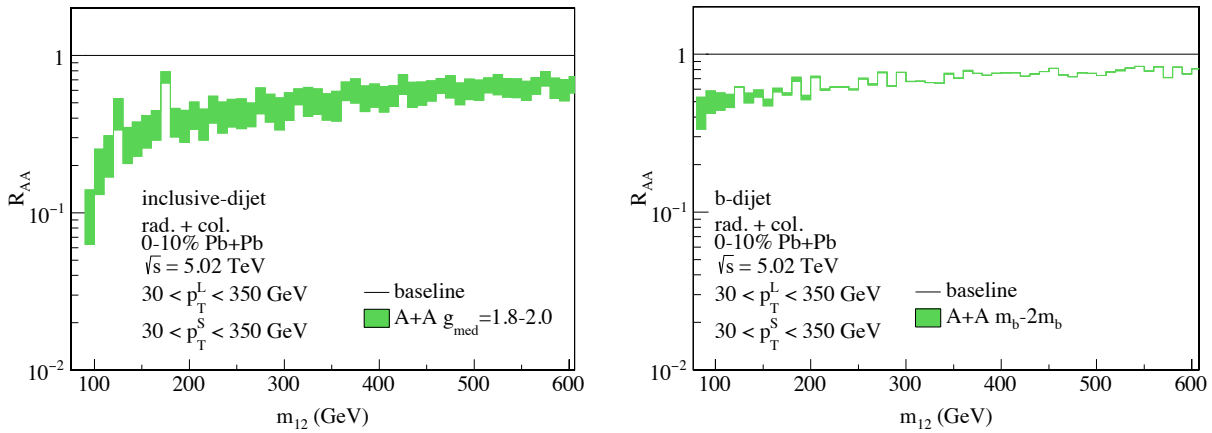

Figure 2.12: Nuclear modification factor R_{AA} is plotted as a function of dijet invariant mass m_{12} for inclusive (left) and b-tagged (right) dijet production in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC. Left: the band corresponds to a range of coupling strength between the jet and the medium: $g_{\text{med}} = 1.8 - 2.0$, respectively. Right: we fix $g_{\text{med}} = 1.8$, and the band corresponds to a range of masses of the propagating system between m_b and $2m_b$.

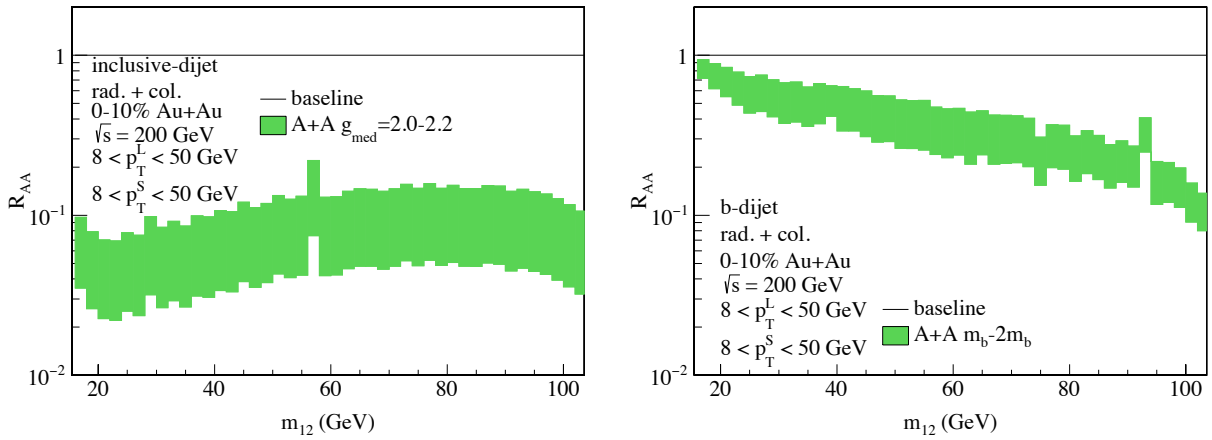

Figure 2.13: Nuclear modification factor R_{AA} plotted as a function of dijet invariant mass m_{12} for inclusive (left) and b-tagged (right) dijet production in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV for sPHENIX at RHIC. Left: the band corresponds to a range of coupling strength between the jet and the medium: $g_{\text{med}} = 2.0 - 2.2$, respectively. Right: we fix $g_{\text{med}} = 2.0$, and the band corresponds to a range of masses of the propagating system between m_b and $2m_b$.

tally. Second, since the dijet invariant mass is proportional to the product of the dijet transverse momenta, as can be clearly seen in Eq. (2.5), the dijet mass distribution incorporates the medium modification of the $d\sigma/d_{1T}dp_{2T}$ in an amplified way, as emphasized in Sec. 2.3.1. In other words, compared to the traditional momentum asymmetry observables, the dijet mass distribution *combines* rather than subtracts the medium modifications of the two jets. Naturally, one would expect the medium modification of dijet mass distributions to be greatly enhanced and thus to be more sensitive to the properties of the medium.

In Fig. 2.12, we plot the nuclear modification factor R_{AA} as a function of dijet invariant mass m_{12} for inclusive (left) and b-tagged (right) dijet production in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC. For inclusive dijet production, the band corresponds to a range of coupling strengths between the jet and the medium: $g_{\rm med} = 1.8 - 2.0$. On the other hand, for b-tagged dijet production, we fix $g_{\rm med} = 1.8$, and the band corresponds to a range of masses of the propagating system between m_b and $2m_b$, implemented as detailed in [HKV13a]. We make transverse momentum cuts requiring both leading and subleading jets to have $p_T^{{\rm L},S} > 30$ GeV. This is why we have a lower limit on the dijet invariant mass $m_{12} \gtrsim 100$ GeV in these plots. As one can clearly see from the figures, being an amplifying effect, R_{AA} can be as small as 0.1, i.e., suppressed by a factor of 10 in the lower end of the invariant mass $m_{12} \sim 100$ GeV. This is a dramatic suppression, much stronger than the suppression for single inclusive jet production, around a factor of 2 [KRV17b]. As one increases the invariant mass m_{12}, the suppression gets smaller, but it is still around a factor of 2 or more even at $m_{12} \sim 500$ GeV. The suppression for b-tagged dijet production is smaller than that of inclusive dijets at smaller dijet mass $m_{12} \sim 100$ GeV, and becomes similar to inclusive dijet production as m_{12} increases. This is to be expected, as heavy quark mass effects on jet quenching are more important at lower transverse momenta, or naturally smaller dijet invariant mass.

In Fig. 2.13, we present the same plots but for Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV, relevant to the sPHENIX experiment at RHIC. For inclusive dijet production, the band corresponds to a range of coupling strengths between the jet and the medium: $g_{\rm med} = 2.0 - 2.2$. On the other hand, for b-tagged dijet production, we fix $g_{\rm med} = 2.0$, and the band again corresponds to a range of masses of the propagating system between m_b and $2m_b$. We choose a slightly larger coupling strength at RHIC compared to that for the above LHC kinematics, which is also consistent with

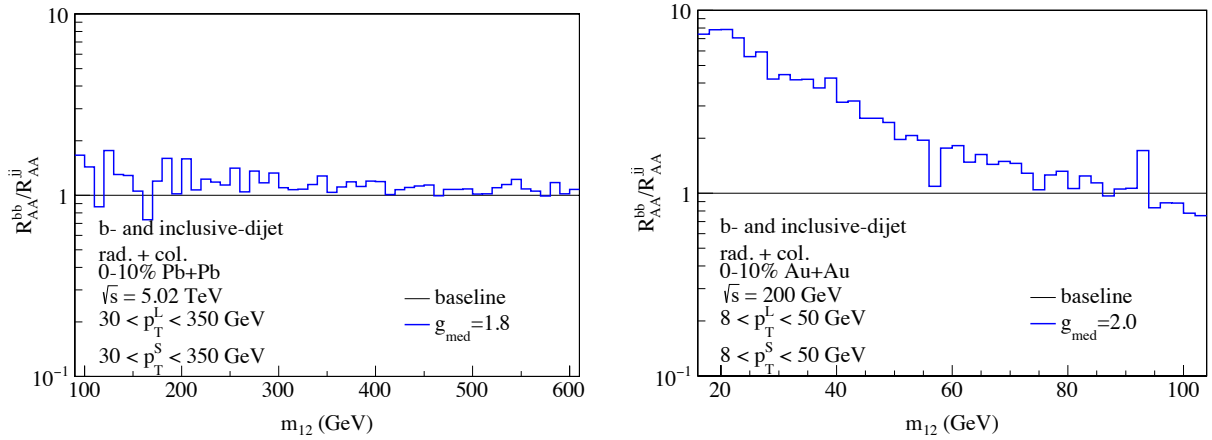

Figure 2.14: Ratios of nuclear modification factors for b-tagged (R_{AA}^{bb}) v.s inclusive (R_{AA}^{jj}) dijet production for CMS (left) and sPHENIX (right) are plotted as a function of dijet invariant mass m_{12}. For LHC (sPHENIX) energies, we choose $g_{\rm med} = 1.8\ (2.0)$. For b-tagged dijets, the mass of the propagating system is held fixed at m_b.

our previous studies and that of the JET collaboration [Bur14]. Since the center-of-mass energy is much lower, we select jets with much lower $p_T \gtrsim 8$ GeV, and correspondingly lower dijet invariant mass $m_{12} \gtrsim 20$ GeV for RHIC kinematics. Having smaller jet transverse momenta and cross sections that fall off strongly as functions of jet transverse momenta, the suppression for inclusive dijet cross sections is even larger compared with those of LHC energies. We observe a factor of ~ 10 or more suppression even up to a relatively high invariant mass $m_{12} \sim 100$ GeV.

On the other hand, the suppression pattern for b-tagged dijet production as a function of m_{12} at sPHENIX energy $\sqrt{s_{NN}} = 200$ GeV, as shown in right panel of Fig. 2.13, appears quite different from inclusive dijet production in left panel, and looks nothing like the b-tagged dijet production at the LHC energy in Fig. 2.12. It is, thus, important to understand why we observe such a behavior. If one recalls the behavior of the suppression pattern for single inclusive heavy meson/heavy quark production as a function of its transverse momentum, see, e.g. Ref. [KRV17a, Cao18], one can understand the above behavior of R_{AA} as a function of m_{12}. Due to the heavy quark mass effect in the jet quenching formalism, R_{AA} for heavy quark mesons first decreases and then increases when plotted as a function of p_T. In other words, there is a dip in R_{AA} as a function of p_T. Now one can translate such a behavior into the behavior of R_{AA} as a function of m_{12}. For the mass region in

Fig. 2.13, b-tagged dijets mostly fall into the relatively low values of jet transverse momenta, i.e., before the dip of R_{AA} (as a function of p_T). This explains why R_{AA} decreases as a function of m_{12}. If one has a larger phase space to explore much higher values of transverse momenta, as is the case at the LHC energy in Fig. 2.12, once passing the dip of R_{AA}, one should naturally expect R_{AA} to increase as a function of m_{12}. This is precisely what is observed in our calculations, see Fig. 2.12 (right). This comparison informs us that sPHENIX is sitting in a very interesting kinematic regime for testing heavy quark mass effects within the jet quenching formalism.

To quantitatively compare the medium modification of b-tagged and inclusive dijet production, we further plot the ratio of nuclear modification factors for b-tagged (R_{AA}^{bb}) and inclusive dijet (R_{AA}^{jj}) production, R_{AA}^{bb}/R_{AA}^{jj}, as a function of dijet invariant mass m_{12} in Fig. 2.14. The left panel shows the results for central Pb+Pb collisions at LHC energy $\sqrt{s_{NN}} = 5.02$ TeV, while the right panel shows the results for central Au+Au collisions at sPHENIX energy $\sqrt{s_{NN}} = 200$ GeV. For LHC (sPHENIX) energies, we choose $g_{\rm med} = 1.8\ (2.0)$. For b-tagged dijets, the mass of the propagating system is held fixed at m_b. In both kinematic regimes, we see a smaller suppression (thus larger R_{AA}) for b-tagged dijets compared to inclusive dijets, though the figure also indicates a markedly different effect at low energies than at higher ones. The most pronounced differences occur in the low mass range $m_{12} \sim 20$ GeV accessible by sPHENIX, where such a ratio reaches up to almost a factor of 10, $R_{AA}^{bb}/R_{AA}^{jj} \sim 10$. On the other hand, at LHC energies, one should observe roughly a factor of 2 less suppression for b-tagged dijet at relatively low dijet invariant mass m_{12}. For large $m_{12} \sim 500$ GeV, the difference diminishes and one should expect to see similar suppressions, $R_{AA}^{bb}/R_{AA}^{jj} \sim 1$.

Let us now turn to the conventional observable, the momentum imbalance distributions, $d\sigma/dz_J$. In the absence of in-medium interactions, one expects from perturbative QCD that the transverse momenta of the two jets are balanced, $p_{1T} \approx p_{2T}$. Consequently, $d\sigma/dz_J$ in elementary p+p collisions will be peaked around $z_J \approx 1$. On the other hand, in heavy ion collisions, jet quenching plays an important role and one jet will lose more energy than the other one. As a result, one expects to see a downshift of the peak in z_J distribution because of strong in-medium interactions.

In Fig. 2.15, we display the normalized dijet imbalance z_J distributions for inclusive (left) and b-tagged (right) dijet production at the LHC energy $\sqrt{s_{NN}} = 5.02$ TeV. The black histogram is

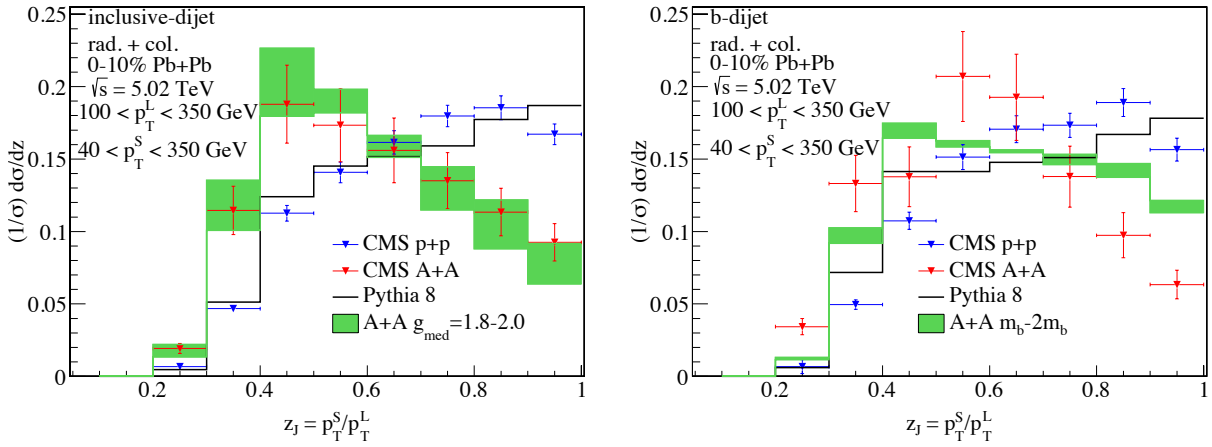

Figure 2.15: The dijet imbalance z_J distributions for inclusive (left) and b-tagged (right) dijet production at $\sqrt{s_{NN}} = 5.02$ TeV for CMS at the LHC. The black histogram is the result for p+p collisions, while the colored curves are the results for central $(0-10\%)$ Pb+Pb collisions. Left: band corresponds to a range of coupling strengths between the jet and the medium: $g_{\rm med} = 1.8-2.0$, respectively. Right: we fix $g_{\rm med} = 1.8$, and the band corresponds to a range of masses of the propagating system between m_b and $2m_b$. The experimental data are from CMS collaboration [Sir18].

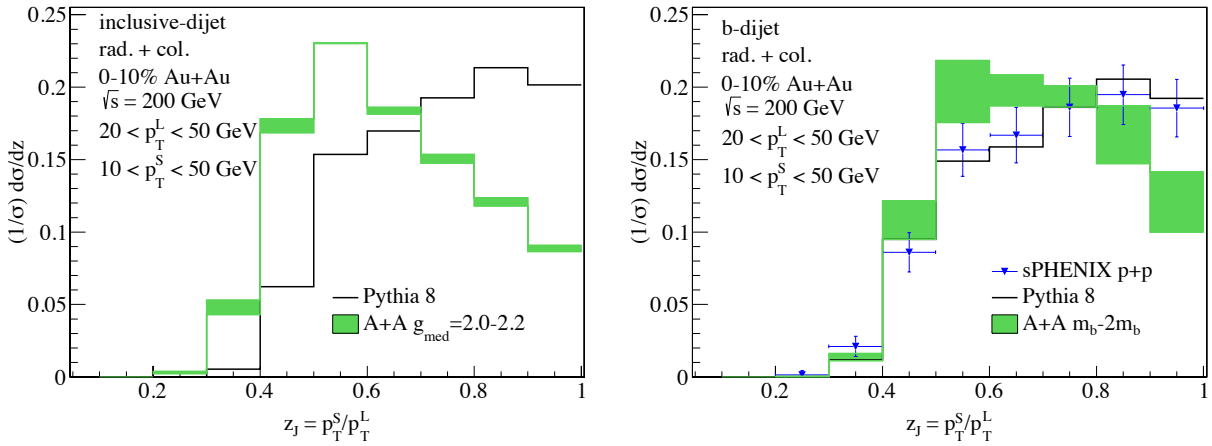

Figure 2.16: The dijet imbalance z_J distributions for inclusive (left) and b-tagged (right) dijet production at $\sqrt{s_{NN}} = 200$ GeV for sPHENIX at RHIC. The black histogram is the result for p+p collisions, while the colored curves are the results for central $(0-10\%)$ Au+Au collisions. The blue "data" points are from preliminary simulations carried out by the sPHENIX collaboration [sPH]. Left: band corresponds to a range of coupling strengths between the jet and the medium: $g_{\rm med} = 2.0-2.2$, respectively. Right: we fix $g_{\rm med} = 2.0$, and the band corresponds to a range of masses of the propagating system between m_b and $2m_b$.

the result for p+p collisions, while the colored curves are the results for central $(0-10\%)$ Pb+Pb collisions. In the left panel, the band corresponds to a range of coupling strengths between the jet and the medium: $g_{\rm med} = 1.8 - 2.0$, respectively. In the right panel, we fix $g_{\rm med} = 1.8$, and the band corresponds to a range of masses of the propagating system between m_b and $2m_b$. The experimental data points are from CMS collaboration [Sir18]. We clearly see a downshift in the peak of z_J distribution for both inclusive and b-tagged dijet production. There is an excellent agreement between our calculations for inclusive dijets and the CMS data. On the other hand, our calculations do not describe very well the CMS data for b-tagged dijets. We attribute this to the use of purely LO matrix elements via Pythia 8 and the specific nature of the re-weighting procedure carried out by CMS [Sir18]. We do not carry out such a re-weighting procedure in order to maintain consistency with the rest of our simulations. Note that the visual difference between our results in A+A and the experimental data is also largely driven by the p+p baseline. Our calculation with $g_{\rm med} = 2.0$ appears closer to the Pb+Pb data. However, as shown below, the results with $g_{\rm med} = 1.8$ already quantitatively capture the downshift of the z_J distribution in heavy ion collisions. This again emphasizes the fact that from the momentum imbalance distributions alone it might be difficult to assess whether a theoretical model correctly represents the physics of jet quenching. Fig. 2.16 contains the dijet imbalance z_J distributions for inclusive (left) and b-tagged (right) dijet production with sPHENIX kinematics at $\sqrt{s_{NN}} = 200$ GeV. Our results for b-tagged dijets in p+p collisions are consistent with the preliminary simulations carried out by the sPHENIX collaboration [sPH] (denoted as the blue "data" points). Our calculations show that a larger shift in z_J should be observed for inclusive dijets compared with b-tagged dijets.

To further quantify the downshift of the z_J distribution, we define the mean value of z_J,

$$\langle z_J \rangle = \left(\int dz_J \, z_J \frac{d\sigma}{dz_J} \right) \Big/ \left(\int dz_J \frac{d\sigma}{dz_J} \right). \tag{2.14}$$

We further define the difference for $\langle z_J \rangle$ in p+p and A+A collisions as

$$\Delta \langle z_J \rangle = \langle z_J \rangle_{\rm pp} - \langle z_J \rangle_{\rm AA}, \tag{2.15}$$

and the positive values of $\Delta \langle z_J \rangle$ represents downshifts of the z_J distribution in A+A collisions in comparison with that of the p+p collisions. In Table 2.1, we list our theoretical calculations for

Table 2.1: Theoretical results for the difference of the average dijet imbalance z_J between p+p and Pb+Pb collisions at $0-10\%$ centrality (CMS) and Au+Au collisions at $0-10\%$ centrality (sPHENIX). Results for CMS may be compared to the experimentally measured values. For both kinematics, we observe a larger shift in imbalance for light flavor dijets than for their heavy counterparts. Both inclusive and b-tagged ranges correspond to the values obtained by varying the coupling to the medium. For CMS: $g_{\rm med} = 1.8 - 2.0$. For sPHENIX: $g_{\rm med} = 2.0 - 2.2$, where the mass of the propagating system is held fixed at m_b.

Kinematics	dijet flavor	$\langle z_J \rangle_{\rm pp}$	$\langle z_J \rangle_{\rm AA}$	$\Delta\langle z_J \rangle$
CMS [Sir18]	b-tagged	0.661 ± 0.003	0.601 ± 0.023	0.060 ± 0.025
Experiment	inclusive	0.669 ± 0.002	0.617 ± 0.027	0.052 ± 0.024
LHC	b-tagged	0.685	0.626 ± 0.013	0.059 ± 0.013
theory	inclusive	0.701	0.605 ± 0.022	0.096 ± 0.022
sPHENIX	b-tagged	0.730	0.665 ± 0.012	0.065 ± 0.012
theory	inclusive	0.743	0.643 ± 0.005	0.100 ± 0.005

$\langle z_J \rangle_{\rm pp}$, $\langle z_J \rangle_{\rm AA}$, and $\Delta\langle z_J \rangle$ for both inclusive and b-tagged dijet production. The values labelled as "LHC theory" are our theoretical calculations for Pb+Pb collisions at $0-10\%$ centrality at $\sqrt{s_{NN}} = 5.02$ TeV, and can be compared with the CMS experimental data. For inclusive dijets, we perform the calculations for the coupling between the jet and the medium $g_{\rm med} = 1.8 - 2.0$, which explains the uncertainties in our theoretical values. For b-tagged dijets, we vary such a coupling in the same range while the mass of the propagating system is held fixed at m_b. We find that in general the downshift $\Delta\langle z_J \rangle$ is slightly larger for inclusive dijet production than that for b-tagged dijets, though the uncertainties are still large. Nevertheless, within the theoretical and experimental uncertainties, our theoretical calculations for all these observables $\langle z_J \rangle_{\rm pp}$, $\langle z_J \rangle_{\rm AA}$, and $\Delta\langle z_J \rangle$, agree well with the CMS experimental data. Finally, we also perform calculations for central Au+Au collisions for sPHENIX kinematics at $\sqrt{s_{NN}} = 200$ GeV in Table 2.1, which are labelled as "sPHENIX theory." We expect such measurements will become available once the sPHENIX experiment starts running in the future.

2.5 Conclusions

In this chapter we present detailed theoretical predictions for inclusive and b-tagged dijet production and modification in heavy ion collisions at RHIC and the LHC. We propose a new observable, the modification of dijet invariant mass, as a novel diagnostic of the QGP created in ultra-relativistic HIC. Our comprehensive studies conclusively demonstrate that this observable exhibits enhanced sensitivity to the strength of jet-medium interactions, the transport properties of nuclear matter, and to the mass effects on in-medium parton showers. This is in sharp contrast to the traditional momentum imbalance measurements, which emphasize only small differences in the quenching of leading and subleading jets. The shift in the mean value of the momentum imbalance variable z_J is only on the order of $7-15\%$ In contrast, the dijet mass modification combines the suppression of the individual jets and enhances the observable jet quenching effect by up to an order of magnitude.

For the main result of this chapter, the dijet mass distribution modification, we find that the suppression at $m_{12} \sim 100$ GeV is around a factor of 10. In contrast, the suppression of single inclusive

jets is only around a factor of 2. In anticipation of experimental measurements from sPHENIX at RHIC, we also perform calculations of dijet mass distributions and momentum imbalance distribution. We find that jet quenching effects on the dijet mass distribution can be significantly amplified in the kinematic range accessible by the future sPHENIX experiment, because of steeply falling spectra. In the mass region $m_{12} = 20 - 100$ GeV, the QGP-induced suppression is a factor of 10 or larger for inclusive dijet production. On the other hand, the suppression for b-tagged dijets shows a different behavior, which can be traced back to the heavy quark mass effects. In other words, at sPHENIX kinematics, there is an enhanced sensitivity to heavy quark mass effects, and we find that in the smaller dijet mass range the suppression for b-tagged dijets can be an order of magnitude smaller.

To conclude, upcoming runs at RHIC and the LHC present compelling opportunities for experiments to explore novel jet quenching observables. The modification of light and heavy flavor dijet mass distributions will be a promising avenue of exploration in this direction.

CHAPTER 3

Heavy flavor dijet production in deep-inelastic scattering as a probe of the gluon Sivers function

---

In this chapter, we demonstrate how the measurement of back-to-back heavy-flavored dijet pairs can provide access to the gluon Sivers asymmetry as manifested in the initial state of deep inelastic scattering experiments. This is done through the development of a transverse-momentum-dependent factorization formalism for heavy flavor dijet production in polarized-proton-electron collisions, which allows us to predict the resulting asymmetry for both charm- and bottom-flavored dijets produced at the future Electron-Ion Collider. Our predictions show that the inclusion of a non-zero quark masses provide important contributions to the overall asymmetry.

3.1 Introduction

In recent years, one of the most important forefronts of hadron physics has been the exploration of the three-dimensional (3D) partonic structure of nucleons in momentum space. Such 3D information is encoded in the so-called transverse-momentum-dependent parton distribution functions (TMD PDFs), which can further inform us about the confined motion of partons in the nucleon, as well as the correlation between their spins, momenta, and the spin of the nucleon [Acc16]. Thanks to semi-inclusive deep inelastic scattering (SIDIS), a great deal of progress has been made in probing and extracting the TMD PDFs of quarks–however, information regarding those of gluons is still largely unknown experimentally. Exploring and measuring gluon TMD PDFs is one of the

primary goals for the future Electron Ion Collider (EIC).

Among the gluon TMD PDFs, the so-called gluon Sivers function is regarded as one of the "golden measurements" at the future EIC [Acc16]. The gluon Sivers function encapsulates the quantum correlation between the gluon's transverse momentum inside the proton and the spin of the proton, thus providing 3D imaging of the gluon's motion. Quite a few processes have been proposed to probe the gluon Sivers function at the EIC, including heavy quark pair production [BMP16], heavy quarkonium production [Yua08, MR17, RKM18, BBP20, BDM20], and quarkonium-jet production [DMP19], as well as back-to-back dihadron and dijet production [Boe11]. The feasibility of measuring the gluon Sivers function in the above scenarios has been studied in [ZAL18], where the authors use the PYTHIA event generator [SMS06] and the reweighting method of [Air10] to investigate the spin asymmetry. They conclude that dijet production is the most promising channel for probing gluon Sivers effects, where the selection of a sufficiently small-x value suppresses the contribution of the quark channel and the corresponding quark Sivers function. In this chapter, we discuss spin asymmetry in the process of heavy flavor (HF) dijet production, where the contribution of the quark Sivers function is further suppressed compared to that of the light flavor dijet case.

An intriguing feature common to both quark and gluon Sivers functions is that they depend non-trivially on the processes in which they are probed. A well-known example of the process-dependence of the quark Sivers function is its sign change between SIDIS and Drell-Yan processes [BHS02, Col02, BMP03]. Similarly, it has been demonstrated that the gluon Sivers function for the process of back-to-back diphoton production in $p+p$ collisions, $p^{\uparrow}p \to \gamma\gamma X$, carries a sign opposite to that of dijet production in $e+p$ collisions, $ep^{\uparrow} \to e'jjX$: $f_{1T,g}^{\perp\left[ep^{\uparrow}\to e'jjX\right]}(x,k_T) = -f_{1T,g}^{\perp\left[p^{\uparrow}p\to\gamma\gamma X\right]}(x,k_T)$ [BMP16]. In [BMM13], it was demonstrated that the gluon Sivers function in any process can be expressed in terms of two "universal" functions with calculable color coefficients for each partonic subprocess. We briefly discuss such a process-dependence for HF dijet production below. For a comprehensive review on gluon TMD PDFs, see [BLP15, Arb20].

So far, studies of the gluon Sivers function at the EIC are mostly performed within the leading-order (LO) parton model, without considering the impact of QCD evolution. The effects of resummation for back-to-back light flavor dijet production in the unpolarized DIS process have been

investigated in [BDD08], where the authors apply the p_T-weighted recombination scheme [ES93] in defining the jet axis to avoid the theoretical complexity arising from non-global logarithm (NGL) resummation [DS01]. A similar idea is used to study single inclusive jet production in the Breit frame at the EIC in [GSW18, GSW19]. Recently, following the same Soft-Collinear Effective Theory (SCET) framework utilized in [BKL18, LRV19, LRV20, CSW19, KLS20], the TMD factorization formula for light flavor dijet production at the EIC has been derived [CEM20], where the azimuthal-angle-dependent soft function, describing the interaction between two final-state jets through the exchange of low-energy gluons, is analytically calculated at one-loop order. For HF jet production in the kinematic region of comparable jet and heavy quark masses, a new effective theory framework is needed. In this work, we provide such a framework and derive the TMD factorization formula.

In the computation of heavy-flavor dijet production, an important feature emerges at NLO which is not present in LO calculations. At LO the two back-to-back heavy-quark jets are represented purely as two heavy quarks traveling without any $\mathcal{O}(\alpha_s)$ corrections stemming from the emission of a gluon. At NLO, one must consider such corrections where each heavy quark radiates a gluon. In order to maintain the back-to-back phase space configuration, the relative transverse momentum between the heavy quark and gluon then must effectively fall below the resolution scale $p_T^{\rm jet} R$, where $p_T^{\rm jet}$ is the transverse momentum of the jet and R is its radius. Conversely, the heavy quark mass m_Q furthermore provides an effective lower limit that the relative transverse momentum must lie above. Thus the NLO heavy quark jet function introduces a non-trivial interplay between parameters regarding the size of the jet, but more importantly, introduces non-trivial heavy-quark mass dependence. This chapter will focus mostly on this aspect, as it is in line with the overall theme of this thesis—that is, the effects of heavy-quark masses on the final state jets measured in particle collisions. Further computational details regarding the gluonic initial conditions can be found in the original paper [KLS20].

The remainder of this chapter is organized as follows. In Sec. 3.2, we detail the factorization framework required to carry out resummation in the back-to-back region where the transverse momentum imbalance of the HF dijet is small. In Sec. 3.3, we present numerical results for charm and bottom dijet production in both unpolarized and transversely-polarized-proton-electron scattering.

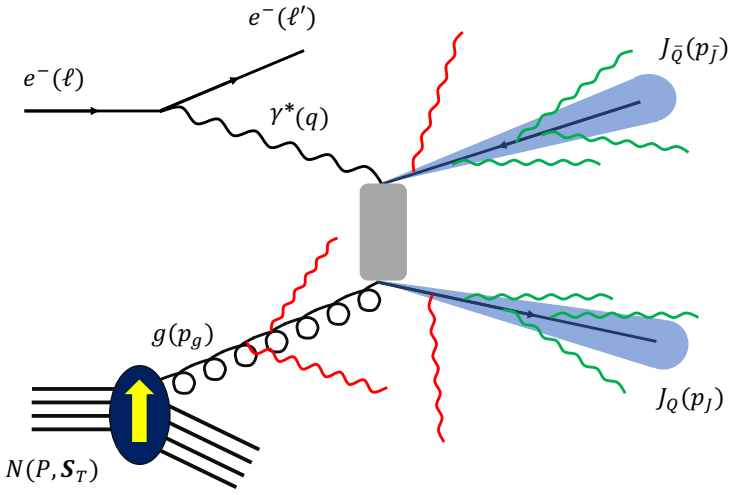

Figure 3.1: HF dijet production in electron-proton collisions, as stated in Eq. (3.1).

We summarize our findings and give an outlook for future investigations in Sec. 3.4.

3.2 Factorization and resummation formula

In this section, we start with the kinematics for HF dijet production in $e + p$ collisions. We then provide the TMD factorization formalism with explicit expressions for all the relevant factorized ingredients.

3.2.1 Kinematics

As shown in Fig. 3.1, we consider HF dijet production in the polarized-proton-electron scattering process

$$e(\ell) + N(P, \boldsymbol{S}_T) \to e(\ell') + J_{\mathcal{Q}}(p_J) + J_{\bar{\mathcal{Q}}}(p_{\overline{J}}) + X\,, \tag{3.1}$$

where $\boldsymbol{S}_T$ is the transverse spin of the polarized proton with momentum P and ℓ (ℓ') is the momentum of the incoming (outgoing) electron. At LO, HF dijets are produced via the $\gamma^* g \to \mathcal{Q}\bar{\mathcal{Q}}$ process. The HF quark $\mathcal{Q}$ and antiquark $\bar{\mathcal{Q}}$ initiate the observed HF jets $J_{\mathcal{Q}}$ and $J_{\bar{\mathcal{Q}}}$ with momentum p_J and $p_{\overline{J}}$, respectively. While the computation of these functions will be the main focus chapter, we still outline the various functions which contribute to the overall factorization of the cross section. In this chapter, we choose to work in the Breit frame so that both the virtual photon (with momentum $q = \ell - \ell'$) and the beam proton scatter along the z-axis. For convenience, we define the following variables commonly used in DIS,

$$Q^2 = -q^2\,, \quad x_B = \frac{Q^2}{2P\cdot q}\,, \quad y = \frac{P\cdot q}{P\cdot \ell}\,. \tag{3.2}$$

We may further note that $Q^2 = x_B\, y\, S_{\ell P}$, where $S_{\ell P} = (\ell + P)^2$ denotes the electron-proton center-of-mass energy. In a fashion analogous to SIDIS, we also define the kinematic variable $z = P\cdot p_J / P\cdot q$, which gives the momentum fraction of the photon carried by the jet $J_{\mathcal{Q}}$. At LO, the four-momenta of the incoming and outgoing particles are expressed as

$$\begin{aligned}
q^\mu &= \frac{Q}{2}(n^\mu - \bar{n}^\mu)\,, \quad P^\mu = \frac{Q}{x_B}\frac{\bar{n}^\mu}{2}\,,\\
\ell^\mu &= \frac{Q}{y}\frac{n^\mu}{2} + Q\frac{1-y}{y}\frac{\bar{n}^\mu}{2} + \ell_t^\mu\,, \quad \ell'^\mu = Q\frac{1-y}{y}\frac{n^\mu}{2} + \frac{Q}{y}\frac{\bar{n}^\mu}{2} + \ell_t^\mu\,,\\
p_J^\mu &= zQ\frac{n^\mu}{2} + \frac{p_T^2}{zQ}\frac{\bar{n}^\mu}{2} + p_t^\mu\,, \quad p_{\overline{J}}^\mu = (1-z)Q\frac{n^\mu}{2} + \frac{p_T^2}{(1-z)Q}\frac{\bar{n}^\mu}{2} - p_t^\mu\,,
\end{aligned} \tag{3.3}$$

where we have introduced two light-like vectors, $n^\mu = (1,0,0,1)$ and $\bar{n}^\mu = (1,0,0,-1)$, and define p_t^μ such that $p_t^\mu p_{t_\mu} = -p_T^2$ with $\boldsymbol{p}_T = p_T(\cos\phi_J, \sin\phi_J)$. We denote transverse momenta relative to the photon-proton beam by the subscript T, while that relative to the jet direction is given the subscript $\perp$. Here, we assume $p_T^2 \gg m_{\mathcal{Q}}^2$ and take $p_J^2 = p_{\overline{J}}^2 = 0$. This allows us to derive the factorized cross section in the following section. Lastly, the parton-level Mandelstam variables

can be defined as

$$\hat{s} \equiv (p_g + q)^2 = (p_J + p_{\bar{J}})^2 = \frac{p_T^2}{z(1-z)}\,, \tag{3.4}$$

$$\hat{t} \equiv (p_g - p_J)^2 = (q - p_{\bar{J}})^2 = -\frac{Q^2 x\, z}{x_B}\,, \tag{3.5}$$

$$\hat{u} \equiv (p_g - p_{\bar{J}})^2 = (q - p_J)^2 = -\frac{Q^2 x\,(1-z)}{x_B}\,, \tag{3.6}$$

where x is the momentum fraction of the proton carried by the gluon, and is given by

$$x = \frac{x_B D}{Q^2 z(1-z)}, \quad \text{with} \quad D = Q^2 z(1-z) + p_T^2\,. \tag{3.7}$$

3.2.2 Factorization formula

In the Breit frame, we define the dijet imbalance as $\boldsymbol{q}_T = \boldsymbol{p}_{JT} + \boldsymbol{p}_{\bar{J}T}$. For this chapter, we examine the back-to-back configuration where $q_T << p_{\bar{J}T} \sim p_{JT} \equiv p_T$. Furthermore, we work in the kinematic regime where $m_{\mathcal{Q}} \lesssim p_T R << p_T$, with R denoting the jet radius. Overall, in the region with the scale hierarchy as $q_T R << q_T \lesssim m_Q \lesssim p_T R << p_T$, the factorized expression for the proton-spin-independent cross section is given by

$$\begin{aligned}\frac{d\sigma^{UU}}{dQ^2 dy d^2\boldsymbol{p}_T dy_J d^2\boldsymbol{q}_T} =& H(Q, y, p_T, y_J, \mu) \int d^2\lambda_T\, d^2k_T\, d^2l_{\mathcal{Q}T}\, d^2l_{\bar{\mathcal{Q}}T} S(\boldsymbol{\lambda}_T, \mu, \nu) \\ &\times \delta^{(2)}(\boldsymbol{\lambda}_T + \boldsymbol{k}_T + \boldsymbol{l}_{\mathcal{Q}T} + \boldsymbol{l}_{\bar{\mathcal{Q}}T} - \boldsymbol{q}_T)\, f_{g/N}\left(x, k_T, \mu, \zeta/\nu^2\right) \\ &\times J_{\mathcal{Q}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\mathcal{Q}}(\boldsymbol{l}_{\mathcal{Q}T}, R, m_{\mathcal{Q}}, \mu)\, J_{\bar{\mathcal{Q}}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\bar{\mathcal{Q}}}(\boldsymbol{l}_{\bar{\mathcal{Q}}T}, R, m_{\mathcal{Q}}, \mu)\,.\end{aligned} \tag{3.8}$$

Above, y_J is the rapidity of the HF jet $J_{\mathcal{Q}}$ and is related to the kinematic variable z through the relation $z = e^{y_J} p_T/Q$. In the factorization formula Eq. (3.8), S denotes the soft function while $f_{g/N}$ is the unpolarized gluon TMD PDF. Their perturbative one-loop expressions can be found in the original paper [KLS20]. In the third line of Eq. (3.8), $J_{\mathcal{Q}}$ and $S^c_{\mathcal{Q}}$ are the massive quark jet and collinear-soft functions, which differ from the corresponding functions utilized in light jet production [BKL18, LRV19, LRV20, CSW19, KLS20]. In Secs. 3.2.3 and 3.2.4, we present their explicit calculations at next-to-leading order (NLO) and highlight the important heavy-quark mass-dependence they introduce. Crucially the above product is required in order to maintain renormalization-group consistency. As we will see, both the HF jet functions and collinear-soft

function will evolve individually according to anomalous dimensions which depend on $m_{\mathcal{Q}}$. The resulting evolution for the entire cross section then becomes perfectly $m_{\mathcal{Q}}$-independent due to the above factorization—this requirement can be understood intuitively by the fact that $m_{\mathcal{Q}}$ is an infrared (IR) scale which must leave the overall ultraviolet (UV) scaling behavior of the cross section invariant. However, as we will. see, the $m_{\mathcal{Q}}$-dependence of the individual jet and collinear soft functions will introduce non-trivial effects. The variables $\boldsymbol{k}_T$, $\boldsymbol{\lambda}_T$, and $\boldsymbol{l}_T$ label the transverse momenta associated with the collinear, soft, and collinear-soft modes. Finally, μ and ν are the factorization and rapidity scales, respectively, while ζ is the Collins-Soper parameter [Col13, ESZ19]. In the derivation of the above factorization formula we apply the narrow jet approximation with $R << 1$. However, as shown in [JSV04, MV12, DDS16, LMR18] this approximation works well even for fat jets with radius $R \sim \mathcal{O}(1)$, and the power corrections of $\mathcal{O}(R^{2n})$ with $n > 0$ can be obtained from the perturbative matching calculation.

Fourier transforming to b-space, the factorized cross section becomes

$$\begin{aligned}\frac{d\sigma^{UU}}{dQ^2 dy d^2\boldsymbol{p}_T dy_J d^2\boldsymbol{q}_T} =& H(Q, y, p_T, y_J, \mu) \int \frac{d^2b}{(2\pi)^2} e^{i\boldsymbol{b}\cdot\boldsymbol{q}_T} S(\boldsymbol{b}, \mu, \nu)\, f_{g/N}\left(x, b, \mu, \zeta/\nu^2\right) \\ &\times J_{\mathcal{Q}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\mathcal{Q}}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)\, J_{\bar{\mathcal{Q}}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\bar{\mathcal{Q}}}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)\,, \end{aligned} \tag{3.9}$$

where soft function S and the gluon TMD PDF $f_{g/N}$ both depend on the rapidity scale ν. Utilizing the relations between this PDF and soft function found in [CJN12, EKM15, Col13, EKM15], Eq. (3.9) can be expressed in the following final form

$$\begin{aligned}\frac{d\sigma^{UU}}{dQ^2 dy d^2\boldsymbol{p}_T dy_J d^2\boldsymbol{q}_T} =& H(Q, y, p_T, y_J, \mu) \int \frac{d^2b}{(2\pi)^2} e^{i\boldsymbol{b}\cdot\boldsymbol{q}_T} S(\boldsymbol{b}, \mu)\, f^{\rm TMD}_{g/N}(x, b, \mu, \zeta) \\ &\times J_{\mathcal{Q}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\mathcal{Q}}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)\, J_{\bar{\mathcal{Q}}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\bar{\mathcal{Q}}}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)\,. \end{aligned} \tag{3.10}$$

Next, if one considers the scattering of an electron with a transversely-polarized proton with spin $\boldsymbol{S}_T$, Eq. (3.8) can be generalized. In this case, the spin-dependent cross section is given by the sum

$$d\sigma(\boldsymbol{S}_T) = d\sigma^{UU} + d\sigma^{UT}(\boldsymbol{S}_T)\,, \tag{3.11}$$

where $d\sigma^{UT}$ depends on the gluon Sivers function. The full expressions for the leading twist gluon distributions are given in [MR01]. Using these results, along with those of [BMM13, BM08], the

factorized cross section then reads

$$\begin{aligned}\frac{d\sigma^{UT}(\boldsymbol{S}_T)}{dQ^2 dy d^2\boldsymbol{p}_T dy_J d^2\boldsymbol{q}_T} =& H^{\text{Sivers}}(Q, y, p_T, y_J, \mu) \int d^2\lambda_T\, d^2k_T\, d^2l_{\mathcal{Q}T}\, d^2l_{\bar{\mathcal{Q}}T} S(\boldsymbol{\lambda}_T, \mu, \nu) \\ &\times \delta^{(2)}(\boldsymbol{\lambda}_T + \boldsymbol{k}_T + \boldsymbol{l}_{\mathcal{Q}T} + \boldsymbol{l}_{\bar{\mathcal{Q}}T} - \boldsymbol{q}_T) \frac{1}{M} \epsilon_{\alpha\beta}\, S_T^\alpha\, k_T^\beta\, f^{\perp,f}_{1T,g/N}\left(x, k_T, \mu, \zeta/\nu^2\right) \\ &\times J_{\mathcal{Q}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\mathcal{Q}}(\boldsymbol{l}_{\mathcal{Q}T}, R, m_{\mathcal{Q}}, \mu)\, J_{\bar{\mathcal{Q}}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\bar{\mathcal{Q}}}(\boldsymbol{l}_{\bar{\mathcal{Q}}T}, R, m_{\mathcal{Q}}, \mu)\,,\end{aligned} \tag{3.12}$$

where H^{Sivers} denotes the hard function for the polarized process, and this expression can once again be written as a Fourier transform by defining

$$\frac{ib^\beta}{2} f^{\perp,f}_{1T,g/N}(x, b, \mu, \zeta/\nu^2) = \int d^2k_T e^{-i\boldsymbol{b}\cdot\boldsymbol{k}_T} \frac{k_T^\beta}{M} f^{\perp,f}_{1T,g/N}(x, k_T, \mu, \zeta/\nu^2)\,. \tag{3.13}$$

Finally, the factorization formula for the polarized differential cross section becomes

$$\begin{aligned}\frac{d\sigma^{UT}(\boldsymbol{S}_T)}{dQ^2 dy d^2\boldsymbol{p}_T dy_J d^2\boldsymbol{q}_T} =& H^{\text{Sivers}}(Q, y, p_T, y_J, \mu) \int \frac{d^2b}{(2\pi)^2} e^{i\boldsymbol{b}\cdot\boldsymbol{q}_T} S(\boldsymbol{b}, \mu) \\ &\times \frac{i}{2}(\epsilon_{\alpha\beta}\, S_T^\alpha\, b^\beta) f^{\perp,f}_{1T,g/N}(x, b, \mu, \zeta) \\ &\times J_{\mathcal{Q}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\mathcal{Q}}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)\, J_{\bar{\mathcal{Q}}}(p_T R, m_{\mathcal{Q}}, \mu)\, S^c_{\bar{\mathcal{Q}}}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)\,.\end{aligned} \tag{3.14}$$

At this point it is useful to take stock of the physics captured by the various components of the above equation. The hard function H^{Sivers} captures the high-energy physics affiliated with the hard scattering of individual partons. It is through this function that the heavy quarks of the final state are tethered to the gluons of the initial state. The soft function S of course describes the ultra-low energy physics of the long-wavelength gluonic radiation which resolves only the net color charge of the incoming and outgoing beams of partons, while the TMD PDF relates to the internal distribution of partons within the incoming proton. This of course leaves us with the HF jet functions, which capture the collinear physics regarding the splitting of heavy quarks into the multi-particle cascades that form the final jets, while the collinear soft function accounts for radiation off heavy quarks of an intermediate energy that enables them to resolve the overall radial extent of the HF jets. All of the heavy-quark dynamics are thus contained within these last two functions, and as such we turn our attention there.

3.2.3 Massive quark jet function

In this section, we discuss the calculation of the massive quark jet function at NLO. The massive quark jet function has been investigated in detail for various observables. For example, the factorization formula for the massive event shape distribution involves such a jet function, as the jet and heavy quark masses are of similar magnitude [FHM08a, FHM08b, LM20, BMP20]. The corresponding jet function has been calculated to two-loop order [HLS19]. Furthermore, the semi-inclusive massive quark jet fragmentation function has been calculated at NLO and applied to inclusive jet production [DKL18, LV19a]. Recently, the one-loop expression for the so-called unmeasured massive quark jet function has been presented in [Kim20].

The global jet anomalous dimension can be obtained from the divergent terms of the unmeasured massive quark jet function. As shown in Fig. 3.2, the one-loop calculation involves two types of diagrams: $J_{\mathcal{Q}}^{\mathrm{NLO,V}}$ and $J_{\mathcal{Q}}^{\mathrm{NLO,R}}$, where $J_{\mathcal{Q}}^{\mathrm{NLO,V}}$ contains only single cut propagators and is thereby unconstrained by the jet algorithm. Explicitly, it is written as

$$J_{\mathcal{Q}}^{\mathrm{NLO,V}} = \frac{\alpha_s}{4\pi} C_F \left[\frac{2}{\epsilon^2} + \frac{1}{\epsilon} \left(1 + 2 \ln \frac{\mu^2}{m_{\mathcal{Q}}^2} \right) + \left(1 + \ln \frac{\mu^2}{m_{\mathcal{Q}}^2} \right) \ln \frac{\mu^2}{m_{\mathcal{Q}}^2} + 4 + \frac{\pi^2}{6} \right] , \tag{3.15}$$

where the heavy quark mass $m_{\mathcal{Q}}$ is the only physical scale involved. Since the real contribution $J_{\mathcal{Q}}^{\mathrm{NLO,R}}$ is constrained by the jet algorithm, it will depend on the jet scale $p_T R$ in addition to $m_{\mathcal{Q}}$. In this work, we define the HF quark four-momentum q^μ with $q^2 = m_{\mathcal{Q}}^2$, which is known as the M-scheme [BMP20]. We note that in the hierarchy of scales we are considering, the constraint of the anti-k_T algorithm [CSS08] is independent of the HF quark mass $m_{\mathcal{Q}}$ and is in fact identical to that for massless partons [DKL18], namely

$$\Theta_{\text{anti-}k_T} = \theta \left[\left(\frac{q^- \left(\omega_J - q^- \right)}{\omega_J} \right)^2 \left(\frac{R}{2 \cosh y_J} \right)^2 - \boldsymbol{q}_\perp^2 \right] , \tag{3.16}$$

where $q^\mu = (q^+, q^-, \boldsymbol{q}_\perp)$ is the four-momentum of the HF quark and ω_J is the large component of the jet four-momentum. The jet scale $p_T R$ emerges in Eq. (3.16) upon noting $\omega_J = 2\, p_T \cosh y_J$. In the phase space integral, we expand the integrated momentum q along the jet direction with

$q^+ = (m_{\mathcal{Q}}^2 + \boldsymbol{q}_\perp^2)/q^-$ given by the power counting requirement $p_T R \sim m_{\mathcal{Q}}$. Explicitly, we have

$$\begin{aligned} J_{\mathcal{Q}}^{\mathrm{NLO,R}}(p_T R, m_{\mathcal{Q}}, \epsilon) &= \frac{\alpha_s C_F e^{\epsilon\gamma_E}\mu^{2\epsilon}}{2\pi\Gamma(1-\epsilon)} \int \frac{dq^-}{\omega_J} \frac{d\boldsymbol{q}_\perp^2}{\boldsymbol{q}_\perp^{2\epsilon}} \left[\frac{q^-}{\omega_J - q^-} \frac{2\boldsymbol{q}_\perp^2 \omega_J^4}{[\boldsymbol{q}_\perp^2 \omega_J^2 + m_{\mathcal{Q}}^2(\omega_J - q^-)^2]^2} \right. \\ &\quad \left. + (1-\epsilon) \frac{\omega_J(\omega_J - q^-)}{\boldsymbol{q}_\perp^2 \omega_J^2 + m_{\mathcal{Q}}^2(\omega_J - q^-)^2} \right] \theta(\omega_J - q^-) \Theta_{\text{anti-}k_T} \\ &= \frac{\alpha_s}{4\pi} C_F \left[-2\ln\left(\frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{m_{\mathcal{Q}}^2} \right) + 2 - \frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} \right] \frac{1}{\epsilon} + J_{\mathcal{Q}}^{\mathrm{R,fin}}, \end{aligned} \tag{3.17}$$

where only a single divergence is exhibited, as the heavy quark mass $m_{\mathcal{Q}}$ acts as a regulator of the overlapping soft and collinear regions of phase space. After combining the real and virtual contributions, the logarithmic dependence on the quark mass $m_{\mathcal{Q}}$ cancels out.

The one-loop global jet renormalization constant then reads

$$Z^{J_{\mathcal{Q}}} = 1 + \frac{\alpha_s}{4\pi} C_F \left[\frac{2}{\epsilon^2} + \frac{1}{\epsilon} \left(2\ln \frac{\mu^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} + 3 - \frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} \right) \right], \tag{3.18}$$

where, again, we observe that the heavy quark mass $m_{\mathcal{Q}}$ only affects the single pole structure. We further note that as $m_{\mathcal{Q}} \to 0$, the massive quark jet renormalization constant reduces to that of the massless jet, $Z^{J_{\mathcal{Q}}} \to Z^{J_q}$. This gives us the following expression for the global jet anomalous dimension

$$\Gamma^{j_{\mathcal{Q}}}(\alpha_s) = -C_F \gamma^{\mathrm{cusp}}(\alpha_s) \ln \frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{\mu^2} + \gamma^{j_{\mathcal{Q}}}(\alpha_s), \tag{3.19}$$

with the one-loop single logarithmic anomalous dimension as

$$\gamma_0^{j_{\mathcal{Q}}} = 2C_F \left(3 - \frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} \right), \tag{3.20}$$

where the first term in the brackets is shared by the massless quark jet function and the second term constitutes the finite quark mass correction. Finally, the renormalized HF jet function is given by the following

$$\begin{aligned} J_{\mathcal{Q},\mathrm{NLO}}^{\mathrm{ren.}}(p_T R, m_{\mathcal{Q}}, \mu) = \frac{\alpha_s}{4\pi} C_F \Bigg[&\left(3 - \frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} \right) \ln \frac{\mu^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} \\ &+ \ln^2 \frac{\mu^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} + 13 - \frac{3\pi^2}{2} + \mathcal{F}(p_T R, m_{\mathcal{Q}}) \Bigg], \end{aligned} \tag{3.21}$$

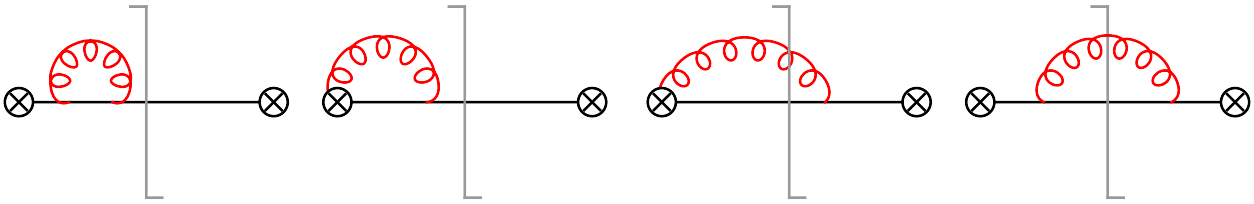

Figure 3.2: Sample Feynman diagrams contributing to the massive quark jet function $J_{\mathcal{Q}}$ at one-loop order in perturbation theory. The virtual corrections $J_{\mathcal{Q}}^{\mathrm{NLO,V}}$ are displayed in the first two diagrams, where each contain only a single cut propagator. The remaining diagrams involving two cut propagators represent the real corrections $J_{\mathcal{Q}}^{\mathrm{NLO,R}}$.

where the function $\mathcal{F}(p_T R, m_{\mathcal{Q}})$ can be expressed as

$$\begin{aligned}\mathcal{F}(p_T R, m_{\mathcal{Q}}) = \pi^2 &- 4\,\mathrm{Li}_2\left(-\frac{m_{\mathcal{Q}}^2}{p_T^2 R^2}\right) + 2\left(1 - \ln\frac{m_{\mathcal{Q}}^2}{p_T^2 R^2}\right)\ln\frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{p_T^2 R^2} \\ &- \frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2}\ln\frac{m_{\mathcal{Q}}^2}{p_T^2 R^2} - \frac{m_{\mathcal{Q}}^2}{p_T^2 R^2}\ln\frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{m_{\mathcal{Q}}^2} \\ &- 4\left[\frac{m_{\mathcal{Q}}}{p_T R}\left(1 + \frac{m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2}\right) + \mathrm{Cot}^{-1}\left(\frac{m_{\mathcal{Q}}}{p_T R}\right)\right]\mathrm{Cot}^{-1}\left(\frac{m_{\mathcal{Q}}}{p_T R}\right). \end{aligned} \tag{3.22}$$

This expression for the HF jet function is equivalent to the semi-analytic form presented in [Kim20], and one can see that as $m_{\mathcal{Q}} \to 0$, we have $\mathcal{F} \to 0$ and, therefore, $J_{\mathcal{Q}} \to J_q$. Hence, the massive quark jet function behaves as expected in the massless limit.

3.2.4 Collinear-soft function

In this section, we calculate the one-loop perturbative expression for the collinear-soft function $S_{\mathcal{Q}}^{c}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \mu)$. The corresponding Feynman diagrams are shown in Fig. 3.3, where the blue and black lines represent Wilson lines along v_J^μ and $\bar{n}_J^\mu$ directions, respectively. Here, the massive quark velocity v_J^μ is defined by

$$v_J^\mu = \frac{\omega_J}{m_{\mathcal{Q}}}\frac{n_J^\mu}{2} + \frac{m_{\mathcal{Q}}}{\omega_J}\frac{\bar{n}_J^\mu}{2}, \qquad \text{with } v_J^2 = 1\,. \tag{3.23}$$

Explicitly, the bare NLO collinear-soft function is given by

$$S_{\mathcal{Q},\mathrm{NLO}}^{c}(\boldsymbol{b}, R, m_{\mathcal{Q}}, \epsilon) = 2C_F\, w_{\bar{n}_J v_J} - C_F\, w_{v_J v_J}\,, \tag{3.24}$$

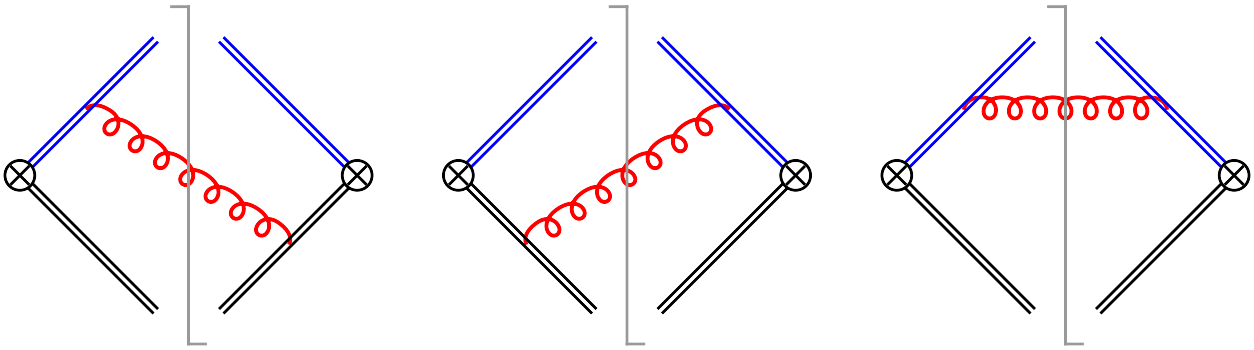

Figure 3.3: One-loop Feynman diagrams of the collinear-soft function $S_{\mathcal{Q}}^c$. The blue and black lines indicate the Wilson lines along v_J^μ and $\bar{n}_J^\mu$ directions, respectively.

where the collinear-soft integrals $w_{\alpha\beta}$ are defined in b-space as

$$w_{\alpha\beta} = \frac{\alpha_s \mu^{2\epsilon} \pi^\epsilon e^{\epsilon\gamma_E}}{2\pi^2} \int d^d k \, \delta^+(k^2) e^{-i\,\bar{n}_J\cdot k\, n_J\cdot b/2} \frac{\alpha\cdot\beta}{(\alpha\cdot k)\,(\beta\cdot k)} \, \theta\left[\frac{n_J\cdot k}{\bar{n}_J\cdot k} - \left(\frac{R}{2\cosh y_J}\right)^2\right] . \tag{3.25}$$

Upon performing the k-integration, we obtain the following expressions for $w_{\alpha\beta}$

$$w_{\bar{n}_J v_J} = \frac{\alpha_s}{4\pi}\left[-\frac{1}{\epsilon^2} - \frac{1}{\epsilon}\left(\ln\frac{\mu^2}{\mu_b^2} + 2\ln\frac{-2ic_{bJ}}{R} - \ln\frac{m_{\mathcal{Q}}^2 + p_T^2R^2}{p_T^2R^2}\right)\right] + w^{\text{fin}}_{\bar{n}_J v_J} , \tag{3.26}$$

$$w_{v_J v_J} = \frac{\alpha_s}{4\pi}\left[-\frac{1}{\epsilon}\left(\frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2R^2}\right)\right] + w^{\text{fin}}_{v_J v_J} . \tag{3.27}$$

We see that the finite quark mass corrections only enter into the single pole structure of the collinear-soft function. This is analogous to the observation made in Sec. 3.2.3 in analyzing the massive quark jet function, and can be understood through the same physical reasoning. The finite terms are given by

$$\begin{aligned} w^{\text{fin}}_{\bar{n}_J v_J} = \frac{\alpha_s}{4\pi}\Bigg[&2\left(\ln\frac{m_{\mathcal{Q}}^2 + p_T^2R^2}{p_T^2R^2} - \ln\frac{-2i\,c_{bJ}}{R}\right)\ln\frac{-2i\,c_{bJ}}{R} \\ &+ \left(\ln\frac{m_{\mathcal{Q}}^2 + p_T^2R^2}{p_T^2R^2} - 2\ln\frac{-2i\,c_{bJ}}{R} - \frac{1}{2}\ln\frac{\mu^2}{\mu_b^2}\right)\ln\frac{\mu^2}{\mu_b^2} + \text{Li}_2\left(-\frac{m_{\mathcal{Q}}^2}{p_T^2R^2}\right) - \frac{\pi^2}{4}\Bigg] , \end{aligned} \tag{3.28}$$

$$w^{\text{fin}}_{v_J v_J} = \frac{\alpha_s}{4\pi}\left[2\ln\frac{m_{\mathcal{Q}}^2 + p_T^2R^2}{p_T^2R^2} - \frac{2m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2R^2}\left(2\ln\frac{-2i\,c_{bJ}}{R} + \ln\frac{\mu^2}{\mu_b^2}\right)\right] . \tag{3.29}$$

Here, we note that $w_{\bar{n}_J v_J}$ reduces to the massless $w_{\bar{n}_J n_J}$ function [BKL18, CSW19] as $m_{\mathcal{Q}} \to 0$, while $w_{v_J v_J}$ vanishes. Given the expression for $S_{\mathcal{Q}}^c$, we can calculate the renormalization constant

$Z^{S^c_{\mathcal{Q}}}$, which is given by

$$Z^{S^c_{\mathcal{Q}}} = 1 + \frac{\alpha_s}{4\pi} C_F \left[-\frac{2}{\epsilon^2} - \frac{2}{\epsilon} \left(\ln \frac{\mu^2}{\mu_b^2} + 2 \ln \frac{-2ic_{bJ}}{R} - \frac{m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} - \ln \frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{p_T^2 R^2} \right) \right] . \tag{3.30}$$

This renormalization constant leads to the following formula for global collinear-soft anomalous dimension

$$\Gamma^{cs_{\mathcal{Q}}}(\alpha_s) = C_F \gamma^{\text{cusp}}(\alpha_s) \ln \frac{R^2 \mu_b^2}{\mu^2} + \gamma^{cs_{\mathcal{Q}}}(\alpha_s) , \tag{3.31}$$

where the one-loop single logarithmic anomalous dimension is

$$\gamma_0^{cs_{\mathcal{Q}}} = -4 C_F \left[2 \ln (-2ic_{bJ}) - \frac{m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} - \ln \frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{p_T^2 R^2} \right] . \tag{3.32}$$

The anomalous dimension for the collinear-soft function associated with the anti-quark is given by

$$\Gamma^{cs_{\bar{\mathcal{Q}}}}(\alpha_s) = \Gamma^{cs_{\mathcal{Q}}}(\alpha_s)|_{\phi_J \to \phi_J + \pi} . \tag{3.33}$$

For phenomenological purposes, we utilize the ϕ_b-averaged collinear-soft function, which can be obtained through Eq. (3.26) by making use of the following integrals

$$\int_0^{2\pi} \frac{d\phi_b}{2\pi} \ln (-2ic_{bJ}) = 0 , \qquad \int_0^{2\pi} \frac{d\phi_b}{2\pi} \ln^2 (-2ic_{bJ}) = -\frac{\pi^2}{6} . \tag{3.34}$$

The resulting anomalous dimension for the ϕ_b-averaged collinear-soft function is denoted by $\bar{\Gamma}^{cs_{\mathcal{Q}}}$ with

$$\bar{\gamma}_0^{cs_{\mathcal{Q}}} = 4 C_F \left(\frac{m_{\mathcal{Q}}^2}{m_{\mathcal{Q}}^2 + p_T^2 R^2} + \ln \frac{m_{\mathcal{Q}}^2 + p_T^2 R^2}{p_T^2 R^2} \right) . \tag{3.35}$$

Upon integrating over ϕ_b, we find that $\bar{\Gamma}^{cs_{\bar{\mathcal{Q}}}}(\alpha_s) = \bar{\Gamma}^{cs_{\mathcal{Q}}}(\alpha_s)$ and, therefore, the two averaged collinear-soft functions $\bar{S}^c_{\mathcal{Q}}$ and $\bar{S}^c_{\bar{\mathcal{Q}}}$ behave identically under QCD evolution.

3.2.5 Renormalization group consistency

Armed with the anomalous dimensions of each component, we are now positioned to demonstrate the RG consistency of our factorization framework.

Inspection of Eqs. (3.19) and (3.31) reveals that all mass corrections cancel exactly in the sum $\Gamma^{j_{\mathcal{Q}}} + \Gamma^{cs_{\mathcal{Q}}}$, making the RG consistency of our formalism identical to the massless case. A similar observation is made in [Kim20]. This general physical behavior has also been observed in the context of inclusive HF jet production [DKL18], where the authors offer the intuitive argument that as the heavy quark mass $m_{\mathcal{Q}}$ constitutes IR information, it thus does not affect the UV behavior of the semi-inclusive jet function. In the present context, we see that the UV evolution behavior of the product of the jet and collinear-soft functions is insensitive to the IR scale introduced by the heavy quark mass. However, in Sec. 3.2.6, we will see how the heavy quark mass enters non-trivially and crucially into the evaluation of the differential cross section.

Therefore, upon combining the anomalous dimensions for the all the functions appearing in our factorized cross sections— of which only those dependent on $m_{\mathcal{Q}}$, i.e. Eqs. (3.18) and (3.31) are explicitly computed in this chapter—the RG consistency of our formalism is established:

$$\Gamma^h + \Gamma^s + \Gamma^{f_g} + 2\Gamma^{j_{\mathcal{Q}}} + \Gamma^{cs_{\mathcal{Q}}} + \Gamma^{cs_{\bar{\mathcal{Q}}}} = 0\,. \tag{3.36}$$

Furthermore, we note that this consistency is preserved under the operation of ϕ_b-averaging

$$\Gamma^h + \bar{\Gamma}^s + \Gamma^{f_g} + 2\Gamma^{j_{\mathcal{Q}}} + \bar{\Gamma}^{cs_{\mathcal{Q}}} + \bar{\Gamma}^{cs_{\bar{\mathcal{Q}}}} = 0\,. \tag{3.37}$$

3.2.6 Resummation formula

Utilizing our EFT framework, all-order resummation is achieved through RG evolution. The resulting all-order expression for the HF dijet production cross section is given at NLL[1] by

$$\begin{aligned}\frac{d\sigma^{UU}}{dQ^2 dy d^2\boldsymbol{q}_T dy_J d^2\boldsymbol{p}_T} &= H(Q, y, p_T, y_J, \mu_h) \int_0^\infty \frac{bdb}{2\pi} J_0(b\, q_T) f_{g/N}(x, \mu_{b*}) \\ &\times \exp\left[-\int_{\mu_{b*}}^{\mu_h} \frac{d\mu}{\mu} \Gamma^h(\alpha_s) - 2\int_{\mu_{b*}}^{\mu_j} \frac{d\mu}{\mu} \Gamma^{j_{\mathcal{Q}}}(\alpha_s) - \int_{\mu_{b_*}}^{\mu_{cs}} \frac{d\mu}{\mu} \left(\bar{\Gamma}^{cs_{\mathcal{Q}}}(\alpha_s) + \bar{\Gamma}^{cs_{\bar{\mathcal{Q}}}}(\alpha_s) \right) \right] \\ &\times \exp\left[-S_{\rm NP}(b, Q_0, n\cdot p_g) \right]\,, \end{aligned} \tag{3.38}$$

[1] In our framework, we ignore contributions from NGL resummation. Such resummation could be included multiplicatively by using the parton shower algorithm developed recently for massive particles [BBF20]. Note that the fitting function used in [CSW19, KLS20] to capture the effects of NGLs is only an approximation for HF jet production, as finite heavy quark mass corrections are not included.

where J_0 is the zeroth order Bessel function of the first kind. In this expression, μ_h, μ_j, and μ_{cs} are the hard, jet, and collinear-soft scales, respectively. We have also performed the usual operator product expansion (OPE) of the unpolarized gluon TMD PDF $f_{g/N}(x, b, \mu, \zeta)$ in terms of the collinear gluon PDF $f_{g/N}(x, \mu)$ at the initial scales $\zeta_i = \mu_i^2 = \mu_{b*}^2$, and have kept the coefficient at LO to be consistent with NLL accuracy. The matching coefficient at higher-orders can be found in e.g. [EKM15, CG12, GLY14, LOS16, ESV16, LYZ20, EMV20]. The function $S_{\rm NP}$ parameterizes the contribution from non-perturbative power corrections which are enhanced for $q_T \sim \Lambda_{\rm QCD}$. Explicitly, we apply the formula given in [SIY18], which reads

$$S_{\rm NP}\left(b, Q_0, n\cdot p_g\right) = g_1 b^2 + \frac{g_2}{2}\frac{C_A}{C_F}\ln\frac{n\cdot p_g}{Q_0}\ln\frac{b}{b_*}\,. \tag{3.39}$$

We assign the following values to the parameters: $g_1 = 0.106\,\mathrm{GeV}^2$, $g_2 = 0.84$ and $Q_0^2 = 2.4\,\mathrm{GeV}^2$, and use the standard b_*-prescription [CSS85]

$$b_* = \frac{b}{\sqrt{1 + b^2/b_{\rm max}^2}}, \qquad \text{with } \; b_{\rm max} = 1.5\ \mathrm{GeV}^{-1} \tag{3.40}$$

in order to regularize the Landau singularity as $b \to \infty$.

Moreover, the spin-dependent cross section is expressed as

$$\begin{aligned}\frac{d\sigma^{UT}(\boldsymbol{S}_T)}{dQ^2 dy d^2\boldsymbol{q}_T dy_J d^2\boldsymbol{p}_T} &= \sin(\phi_q - \phi_s)\, H(Q, y, p_T, y_J, \mu_h) \int_0^\infty \frac{b^2 db}{4\pi} J_1(b\, q_T) f_{1T,g/N}^{\perp,f}(x, \mu_{b*}) \\ &\times \exp\left[-\int_{\mu_{b*}}^{\mu_h}\frac{d\mu}{\mu}\Gamma^h\left(\alpha_s\right) - 2\int_{\mu_{b*}}^{\mu_j}\frac{d\mu}{\mu}\Gamma^{j_{\mathcal{Q}}}\left(\alpha_s\right) - \int_{\mu_{b_*}}^{\mu_{cs}}\frac{d\mu}{\mu}\left(\bar{\Gamma}^{cs_{\mathcal{Q}}}\left(\alpha_s\right) + \bar{\Gamma}^{cs_{\bar{\mathcal{Q}}}}\left(\alpha_s\right)\right)\right] \\ &\times \exp\left[-S_{\rm NP}^{\perp}(b, Q_0, n\cdot p_g)\right]\,. \end{aligned} \tag{3.41}$$

Here, we have expected a similar OPE for the gluon Sivers function $f_{1T,g/N}^{\perp}(x, b, \mu, \zeta)$ at the initial scales $\zeta_i = \mu_i^2 = \mu_{b*}^2$ and simply expressed the corresponding collinear function at LO as $f_{1T,g/N}^{\perp,f}(x, \mu_{b*})$ for simplicity. In principle, the corresponding collinear functions in the OPE expansion would be the twist-3 three-gluon correlation functions defined in [Ji92, BKT10]. To the best of our knowledge, detailed OPE calculations for the corresponding coefficient functions are not available in the literature. An expansion of the gluon Sivers function in terms of the collinear twist-3 quark-gluon-quark correlator, or the so-called Qiu-Sterman function [QS91, QS92], in transverse momentum space is performed in [Yua08]. On the other hand, the coefficient functions for the expansion of the quark Sivers function in terms of the three-gluon correlation functions are provided

in [DKP15, STV19]. The computation of the coefficient functions for expanding the gluon Sivers function in terms of the three-gluon correlation functions is essential for a full understanding of the QCD evolution of the gluon Sivers function. We leave this to future work.

Our knowledge about gluon Sivers functions, especially in the proper TMD factorization formalism, is rather limited. At the present moment, the only experimental constraint on the gluon Sivers function, in the TMD framework, comes from the SIDIS measurement of back-to-back hadron pairs off transversely-polarized deuterons and protons at COMPASS [Ado17]. However, as of yet, there has been no theoretical extraction of the gluon Sivers function from such data. On the other hand, an important theoretical constraint on the gluon Sivers function comes from the Burkardt sum rule [Bur04]. For the phenomenological purposes of the next section, we adopt the non-perturbative parameterization utilized by [DMP15, ADM16] [2]. Specifically, for the non-perturbative Sudakov, we take

$$S_{\rm NP}^{\perp}\left(b, Q_0, n\cdot p_g\right) = g_1\rho\, b^2 + \frac{g_2}{2}\frac{C_A}{C_F}\ln\frac{n\cdot p_g}{Q_0}\ln\frac{b}{b_*}\,, \tag{3.42}$$

where the g_2-dependent term is spin-independent and is, therefore, the same term occurring in Eq. (3.39), while the term $\propto g_1\rho$ can be connected to the Gaussian width in transverse momentum space [EIK14] for the gluon Sivers function. For the collinear part of the gluon Sivers function, $f_{1T,g/p}^{\perp,f}(x,\mu)$ in Eq. (3.41), we take

$$f_{1T,g/N}^{\perp f}(x,\mu) = N_g\frac{4\rho\sqrt{2e\rho(1-\rho)g_1}}{M_{\rm proton}}x^{\alpha_g}(1-x)^{\beta_g}\frac{(\alpha_g+\beta_g)^{\alpha_g+\beta_g}}{\alpha_g^{\alpha_g}\beta_g^{\beta_g}}f_{g/N}(x,\mu)\,, \tag{3.43}$$

with the parameters given by

$$N_g = 0.65,\quad \alpha_g = 2.8,\quad \beta_g = 2.8,\quad \rho = 0.5,\quad M_{\rm proton} = 1\,{\rm GeV}\,, \tag{3.44}$$

and $f_{g/N}(x,\mu)$ denoting the unpolarized collinear gluon PDF. For $f_{g/N}(x,\mu)$, we use CT14nlo [DHG16]–specifically, CT14nlo_NF3 (CT14nlo_NF4) for charm (bottom) jet-pair production with 3 (4) active parton flavors.

[2]Note that the gluon Sivers function in [DMP15], and its updated version [DFM19], is constrained to their study of the $p^{\uparrow}p\to\pi X$ process. Technically, this is not subject to a TMD factorization framework, but it serves as a starting point for our numerical study, following [ZAL18].

At this point, it is important to note that while the mass corrections in sum of the anomalous dimensions for the collinear-soft and massive jet functions cancel, the mass-dependence of Γ^{j_Q} contributes to the differential cross section. By examining Eqs. (3.38) and (3.41), we see that the mass corrections enter into the evolution between the scales μ_j and μ_{cs}. We will see in the following section that this can significantly affect both the q_T-distributions and spin asymmetries for HF dijet production at the EIC.

3.3 Numerical results

In this section, we present numerical results for HF dijet production in unpolarized and transversely-polarized-proton-electron collisions at the future EIC. We set the energies of the electron and proton beam to be 20 GeV and 250 GeV, respectively. These beam-energy values yield a electron-proton center-of-mass energy of $\sqrt{S_{\ell P}} = 141$ GeV. For the all-order resummation formulae in Eqs. (3.38) and (3.41), the renormalization scales for each function are chosen to be

$$\mu_h = \sqrt{Q^2 + p_T^2}, \quad \mu_j = p_T R, \quad \mu_{cs} = \mu_{b_*} R\,. \tag{3.45}$$

Here, note that the Landau singularity associated with the collinear-soft scale is also regularized by the b_*-prescription.

As given in the calculation of the jet function, we consider HF jets constructed using the anti-k_T algorithm with radius $R = 0.6$. The corresponding kinematic cuts for charm and bottom jets in the Breit frame are

$$\begin{aligned} \text{charm jets}: \ & 5\,\text{GeV} < p_T < 10\,\text{GeV}, \quad |y_J| < 4.5\,, \\ \text{bottom jets}: \ & 10\,\text{GeV} < p_T < 15\,\text{GeV}, \quad |y_J| < 4.5\,, \end{aligned} \tag{3.46}$$

respectively. The charm and bottom quark masses are chosen as $m_c = 1.5\,\text{GeV}$ and $m_b = 5\,\text{GeV}$. The spin asymmetry from the gluon Sivers function is defined as

$$A_{UT}^{\sin(\phi_q - \phi_s)} = 2 \frac{\int d\phi_s d\phi_q \sin(\phi_q - \phi_s)\, d\sigma^{UT}(\boldsymbol{S}_T)}{\int d\phi_s d\phi_q\, d\sigma^{UU}}\,. \tag{3.47}$$

In Fig. 3.4, we display the normalized unpolarized cross section, $1/\sigma\, d\sigma/dq_T$, as a function of the imbalance q_T. In Fig. 3.5, the Sivers spin asymmetry $A_{UT}^{\sin(\phi_q - \phi_s)}$ is presented as a function of

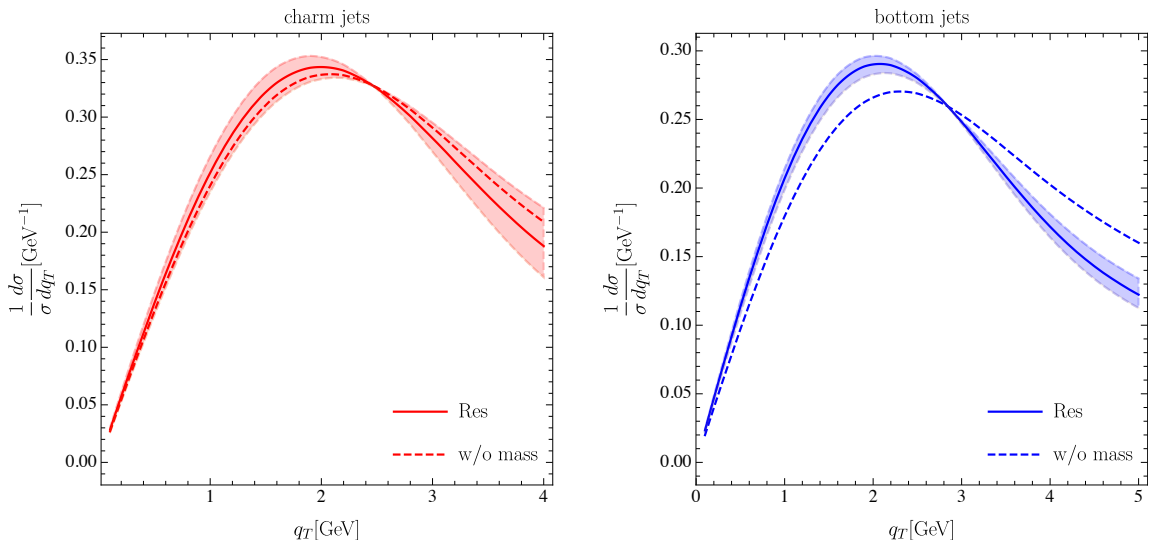

Figure 3.4: The normalized q_T-distribution for the unpolarized cross section of charm (left plot) and bottom (right plot) dijet production at the EIC. The solid curves are the results from using the resummation formula Eq. (3.38), while the dashed curves represent the resummation prediction using the evolution kernel without finite quark mass corrections. The red and blue bands indicate theoretical uncertainties from the variation of hard and jet scales as discussed in the text.

q_T/p_T following [AKP20], for both charm (left panel) and bottom (right panel) jets, respectively. For both plots, the solid curves are the results obtained using the resummation formula, while the dashed curves represent the resummation prediction using the evolution kernel without finite quark mass corrections. For both the unpolarized q_T and $A_{UT}^{\sin(\phi_q-\phi_s)}$ distributions, we find that the effects of the finite quark masses are modest for charm jets and quite sizable for bottom jets. This can be attributed to the sizes of the charm and bottom masses relative to their associated jet scales $p_T R$. As discussed in Secs. 3.2.3 and 3.2.4, we have that $J_{\mathcal{Q}} \to J_q$ and $S^c_{\mathcal{Q}} \to S^c_q$ as $m_{\mathcal{Q}} \to 0$, making them analytic functions of $m_{\mathcal{Q}}$ in the neighborhood of zero mass. Since Eqs. (3.38) and (3.41) carry their mass-dependence through the anomalous dimensions for the jet and collinear-soft functions, Eqs. (3.19) and (3.31), one sees that the massive versions of these functions are connected to the massless versions by the ratio $m_{\mathcal{Q}}/(p_T R)$–it is in fact this dimensionless parameter that controls the physical size of the mass corrections. With this in mind, one sees that Eq. (3.46) naturally positions bottom dijets further (in terms of the parameter $m_{\mathcal{Q}}/(p_T R)$) from light flavor jet-pairs

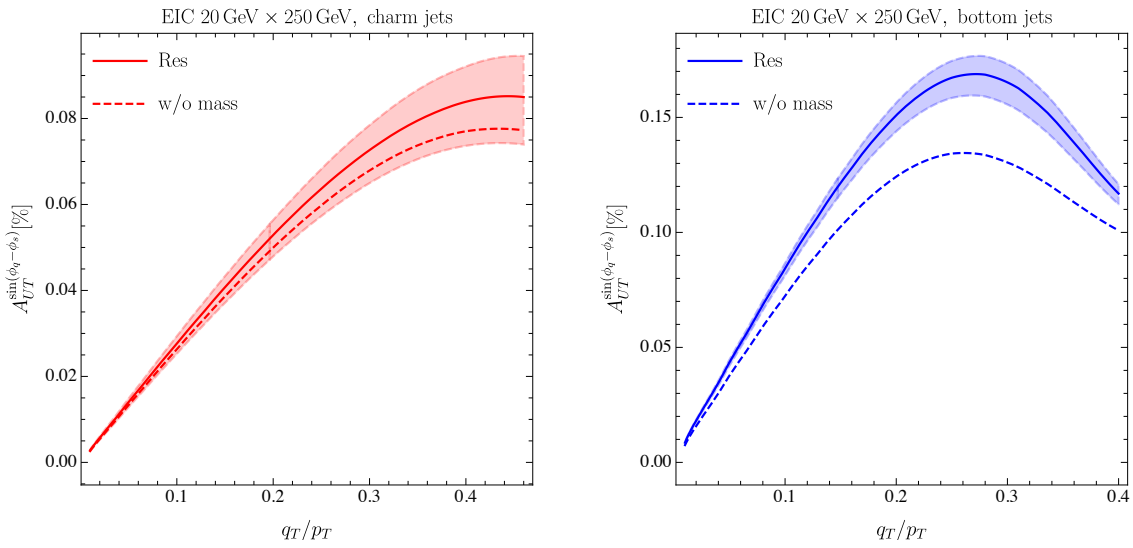

Figure 3.5: The Sivers spin asymmetry for charm (left plot) and bottom (right plot) dijet production at the EIC is plotted as a function of q_T/p_T. The solid curves are the results from using the resummation formula, while the dashed curves represent the resummation prediction using the evolution kernel without finite quark mass corrections. The red and blue bands indicate theoretical uncertainties from the variation of hard and jet scales.

than it does charm dijets. This relative positioning is then clearly displayed in Figs. 3.4 and 3.5.

In order to estimate the theoretical uncertainties, in both Figs. 3.4 and 3.5 we also show the uncertainties from scale variations, which are given by the red and blue bands. Here we vary the hard and jet scales by a factor of two around their default values as defined in (3.45), and the total uncertainty bands are obtained by the envelope of all the variations. Since the non-perturbative Sudakov factor in Eq. (3.42) is fitted at the canonical scale μ_{b_*}, we do not include theory uncertainties from μ_{b_*} and μ_{cs} variations. We find that the scale uncertainty is compatible with the finite quark mass corrections in charm dijet process, while its impact on the bottom dijet process is smaller than the mass correction. Therefore in order to identify the finite quark mass effects in the charm dijet process it is essential to reduce the scale uncertainties. Our factorization and resummation formula provides a clear structure to improve the perturbative accuracy, which makes scale uncertainty further reduction possible. We leave the higher-order perturbative calculations in future work.

3.4 Conclusion

A major priority of the future EIC is to explore the gluon TMD PDFs. In this chapter, we have investigated the use of back-to-back HF dijet production in transversely-polarized target DIS as a means of probing spin-dependent gluon TMD PDFs. We have calculated the expressions for the mass-dependent jet and collinear-soft functions at next-to-leading order. Using these expressions, as well as Soft-Collinear Effective Theory, we resum the large logarithms associated with these expressions at next-to-leading logarithmic accuracy. We then provide a factorization theorem for this process with QCD evolution in the kinematic region where heavy quark mass $m_{\mathcal{Q}} \lesssim p_T R << p_T$, with p_T and R being the transverse momentum and the radius of the jet, respectively. Furthermore, we generate a prediction for the Sivers asymmetry for charm and bottom dijets at the EIC, which can be used to probe the gluon Sivers function. This is due to the fact that, at the level of the hard scattering, the initial state which dominates the production of heavy-quark pairs in the final state involves the gluon from the incoming proton and the virtual photon emitted from the inbound electron. Thus, through this particular dijet configuration measured in the final state, one can infer its microscopic origins and ultimately extract valuable information about the proton. Furthermore, we carefully study the effects of the HF masses by comparing our mass-dependent predicted asymmetry against the asymmetry in the massless limit. We find that, in the kinematic region we consider, the HF masses generate modest corrections to the predicted asymmetry for charm dijet production but sizable corrections for the bottom dijet process. Furthermore, we also consider the theoretical uncertainties from the scale variation. We find that the scale uncertainty can be compatible with the corrections from finite quark mass effects, especially for charm dijets production. In order to identify the mass effects and reduce the scale uncertainties one has to include higher-order corrections in the matching coefficients and the corresponding anomalous dimensions in Eqs. (3.38) and (3.41), and we leave the detailed perturbative calculations in future work.

Part II

Heavy quark jet substructure

CHAPTER 4

The dead-cone angle through cumulants of subjet-energy distributions

---

A striking prediction of quantum chromodynamics is the suppression of gluonic radiation off heavy quarks relative to light quarks and gluons—this is known as the dead-cone effect. In this chapter, we demonstrate how traditional jet substructure observables for jets initiated by heavy quarks manifest the dead-cone effect. In particular, we consider the fragmentation of jet of radius R to a subjet of radius $r < R$, clustered about the Winner-Take-All axis, which is preferentially directed along the trajectory of the initiating heavy quark. We compute the energy distribution of such subjets and show how the dead-cone angle imprints itself on the r-dependence of its first two cumulants. This is understood theoretically to arise from hitherto unexplored finite terms, appearing at next-to-leading-order, involving the heavy-quark mass, transverse momentum of the jet, and the subjet radius r. We present theoretical predictions for the cumulants of this distributions—the first of which is traditionally referred to as the jet shape—for heavy-quark-initiated jets to be measured by sPHENIX at RHIC.

4.1 Introduction

Deciphering the substructure of jets emerging from hadronic and nuclear collisions is a principal means through which modern collider physicists probe the inner-workings of the Standard Model in general and Quantum Chromodynamics (QCD) in particular. The overarching goal of such

investigations is the determination of how physical features of the elementary quarks and gluons generating the jet imprint themselves on the final collection of hadrons reaching calorimeter cells. Of particular interest are jets containing heavy quarks, such as charm and bottom. For instance, the dominant $H \to b\bar{b}$ channel [AAA17, STA18] provides an important constraint on the decay width of the Higgs. Jets containing b-quarks emerging from heavy-ion collisions probe the flavor-dependence of parton energy loss through a nuclear medium [HKV13b, HKV15b, LV19c, LV19b], where the decay patterns of b-dijet pairs have been shown to be particularly sensitive probes of the quark-gluon plasma [DWZ18b, KRV19b]. Bottom dijets furthermore provide access to initial-state aspects of the proton, such as the Sivers function [KRS21].

However, a particularly interesting feature of heavy quarks is the way in which their most distinctive attribute, that is, their mass m_Q, manifests itself within the internal radiation pattern of a jet. The dead-cone effect [DKT91] refers to the suppression of gluon radiation off heavy quarks within an conical region with radius of order m_Q/E_Q, where E_Q is the energy of the heavy quark. This arises due to m_Q effectively providing an infrared (IR) cutoff which shields the collinear divergences characteristic of massless particles. This effect has been measured by the ALICE collaboration [CP19, Zar21, Ach22], albeit through an observable which cannot be computed through the use of perturbative QCD. An alternative observable which both reveals the dead-cone angle and is amendable to perturbative calculation is the energy-energy correlator [CLM22b]. The goal of this work is to provide the computational details a complementary observable through which the dead-cone angle is revealed in the most direct and simplest fashion possible.

We consider the case of a jet of radius R fragmenting to a subjet of radius $r < R$ centered about the Winner-Take-All (WTA) axis [BCT14a, LNT14]. This axis is defined as that directed along the most energetic particle within a jet which, for jets containing heavy quarks, is predominantly the heavy quark itself. Therefore subjets centered about this axis give direct access to the conical region surrounding the heavy quark, of which the dead-cone is a subset. We consider the ratio of the subjet energy to that of the overall jet, denote this fraction as $z_{\rm sub} \equiv p_T^{\rm subjet}/p_T^{\rm jet}$, and compute the probability distribution for these values $p(z_{\rm sub})$. We show that the first two cumulants of this distribution, treated as functions of the subjet radius r, exhibit well-defined peaks at $r \sim m_Q/p_T^{\rm jet}$ when treated differentially, providing clean access to this fundamnental prediction of QCD: the

dead-cone angle.

The first of the aforementioned cumulants corresponds to a traditional jet substructure observable, known as the jet shape, centered about the WTA axis. Calculations of the jet shape for massless progenitors have been carried out for both pp collisions [Sey98, Sey97, CV14a, CRW19a] as well as AA collisions [CV16b, VWZ08b] for jets clustered about a central jet axis as well as about the WTA axis in [KRW17, NPW19a]. The second cumulant corresponds to the variance of $p(z_{\rm sub})$ and as such is a novel observable due to its non-linear relation to the jet shape. Thus, we access the dead-cone angle through the two simplest and intuitive summary statistics of the subjet-energy distribution $p(z_{\rm sub})$.

This chapter is organized as follows: the theoretical formalism for our computation is contained in Sec. 4.2 wherein Sec. 4.2.1 outlines the factorization required for the computation of the subjet-energy distribution and describes the experimental determination of its cumulants. Fixed-order analysis resides in Sec. 4.2.2 and the all-order effects of QCD evolution are contained in Sec. 4.2.3. Our phenomenological predictions for the sPHENIX experiment are presented in Sec. 4.3 and we conclude in Sec. 7.5.

4.2 Formalism

4.2.1 Factorization

Since we are ultimately after the dead-cone angle, we want to measure jets originating from heavy-quarks in the hard scattering, and thus need to measure jets whose dominant channel comes from such a situation. Pythia simulations [Bie22] carried out in [KRV19b], reveal that for sPHENIX kinematics, back-to-back b-dijet pairs are sourced overwhelmingly by hard b-quarks, and thus fit our bill quite effectively. Thus, by first identifying the leading jet in a back-to-back b-dijet pair, and then measuring the subjet-energy distribution on the leading jet, we can extract the distribution in a clean way.

The factorization for such a process takes the following form:

$$\frac{d\sigma}{dz_{\rm sub}\,dp_T} = \frac{d\hat{\sigma}_Q}{dp_T} \times \mathcal{G}_Q^{\rm jet}(z_{\rm sub}, m_Q, p_T, r, R)\,, \tag{4.1}$$

where $d\hat{\sigma}_Q/dp_T$ is the hard-scattering cross-section and $\mathcal{G}_Q^{\rm jet}$ is the exclusive subjet fragmentation function, which differs from the semi-inclusive version as computed in [KRW17] by the omission of collinear radiation falling outside the jet cone, as this would take us outside the back-to-back phase space configuration. The cross section for the production of a Q-initiated jet is obtained by integrating the above cross section over $z_{\rm sub}$ and thus obeys

$$\frac{d\sigma}{dp_T} = \frac{d\hat{\sigma}_Q}{dp_T} \times \mathcal{J}_Q(m_Q, p_T, R)\,, \tag{4.2}$$

where $\mathcal{J}_Q(m_Q, p_T, R)$ is the exclusive jet function, as computed in [Kim20, KRS21]. Thus all initial-state dependence through the hard-scattering cancels in the normalized cross section

$$\begin{aligned}\frac{d\sigma}{dz_{\rm sub}\,dp_T}\bigg/\frac{d\sigma}{dp_T} &= \mathcal{G}_Q^{\rm jet}(z_{\rm sub}, m_Q, p_T, r, R)\big/\mathcal{J}_Q\left(m_Q, p_T, R\right)\,, \\ &\equiv p(z_{\rm sub})\,. \end{aligned} \tag{4.3}$$

Thus the computation of $p(z_{\rm sub})$ amounts to that of the exclusive jet and subjet fragmentation functions initiated by a heavy quark Q.

4.2.2 Subjet function

In this section, we calculate the subjet energy distribution for a jet intitiated by a heavy quark. This object will be a distribution that is differential in the variable $z_{\rm sub}$, which we define to be the ratio of the energy of the subjet (of radius r) to the energy of the overall jet (of radius R). The LO contribution comes simply from a heavy quark propagating freely—a single particle in both the initial and final states. Since the heavy quark is the only particle forming both the subjet and jet, $z_{\rm sub}$ is identically one, and therefore its contribution to the distribution is $\delta(1 - z_{\rm sub})$. We will refer to this configuration as "(LO)." A similar one-particle initial/final state combination arises at NLO from the correction due to the emission and re-absorption of a virtual gluon by the heavy quark.

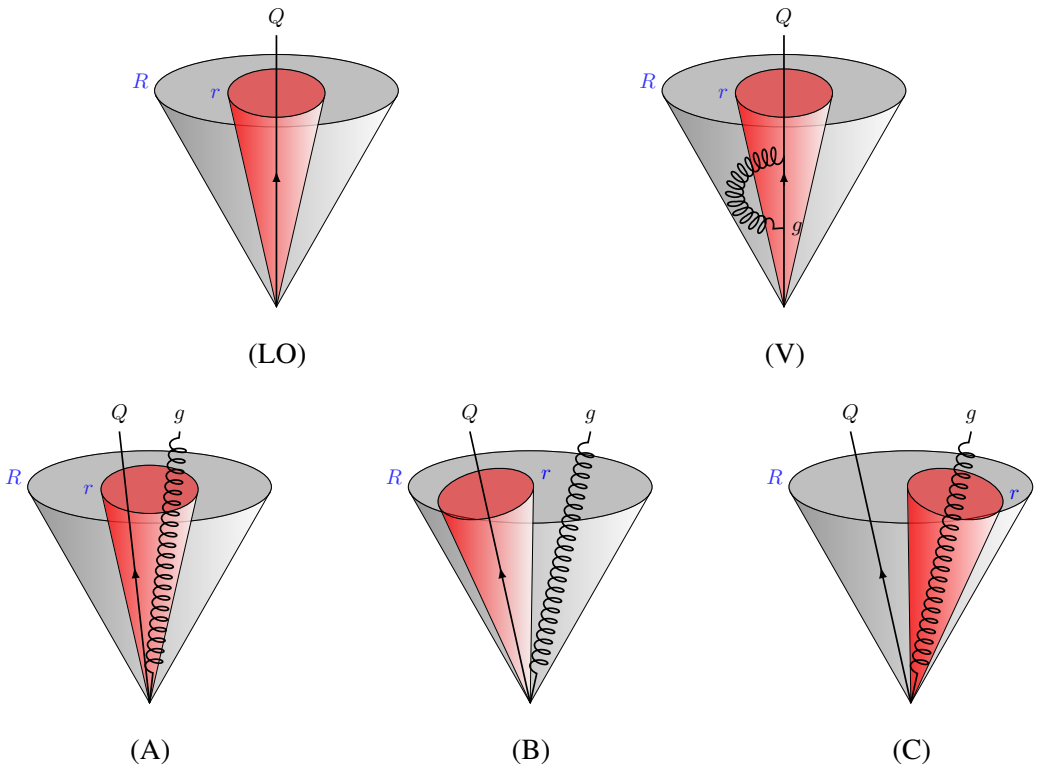

Figure 4.1: The phase space configurations contributing to the subjet function through $\mathcal{O}(\alpha_s)$. (LO) and (V): the heavy quark is the sole particle making up the jet and subjet. (A): both partons are unresolved and fall in the subjet cone. (B): the splitting is resolved and the heavy quark carries more of the energy and hence defines the WTA axis. (C): the splitting is resolved and the gluon defines the WTA axis.

This contribution similarly has delta-function support:

$$\Pi_V\left(z_{\text{sub}}; m_Q\right) = \delta(1 - z_{\text{sub}})\frac{\alpha_s C_F}{2\pi}\left[\frac{1}{\epsilon^2} + \frac{1}{2\epsilon} + \frac{1}{\epsilon}\log\frac{\mu^2}{m_Q^2} + \frac{1}{2}\log^2\frac{\mu^2}{m_Q^2} + \frac{1}{2}\log\frac{\mu^2}{m_Q^2} + \frac{\pi^2}{12} + 2\right]. \tag{4.4}$$

We will refer to this phase space configuration as (V) for "virtual." The notational choice for the arguments going as $(z_{\text{sub}}; \dots)$ is in anticipation of the probabilistic character the subjet function will take on with respect to the random variable z_{sub}. Such notation will be maintained for the real emission contributions, from whence the subjet function acquires functional dependence on p_T, r, and R in addition to m_Q.

The remaining phase space contributions at NLO involve two-particle final states arising from

the real emission of a gluon from the heavy quark. These real emission diagrams take the form of a sum over phase space integrals of the heavy-quark-initiated splitting functions [KRV17a]

$$\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)=\frac{C_F}{\ell_\perp^2+(1-z)^2m_Q^2}\left(\frac{1+z^2}{1-z}-\epsilon\left(1-z\right)-\frac{2m_Q^2\,z(1-z)}{\ell_\perp^2+(1-z)^2m_Q^2}\right), \tag{4.5}$$

$$\hat{P}_{gQ\leftarrow Q}\left(z,\ell_\perp\right)=\frac{C_F}{\ell_\perp^2+z^2m_Q^2}\left(\frac{1+(1-z)^2}{z}-\epsilon\,z-\frac{2m_Q^2\,z(1-z)}{\ell_\perp^2+z^2m_Q^2}\right), \tag{4.6}$$

where z and $\ell_\perp$ are the energy fraction and transverse momentum, respectively, of the daughter parton with respect to the initial heavy quark, and m_Q is the heavy quark mass. The above splitting functions are the central objects of our calculation. The subjet function receives contributions from three separate real emission phase space configurations, which we will label by (A), (B), and (C). These, along with (LO) and (V), are depicted in Fig. 4.1.

In what follows, we will speak in terms of resolved and unresolved splittings. A resolved splitting is one for which the transverse momentum of the daughter parton $\ell_\perp$ is greater than some resolution scale $k_\perp$. In this work, since we are interested in the angular-dependence about the WTA axis, we will express this resolution scale in terms of the angle between the two daughter partons $k_\perp = z(1-z)p_Tr$. This choice amounts to the subjet being clustered according to the anti-k$_T$ constraint. An unresolved splitting is then the case where $\ell_\perp$ is less than this resolution scale.

The first real emission contribution, referred to as (A), is the configuration in which the splitting is unresolved, i.e. both final state partons are contained within the subjet radius r. Thus the entirety of the jet's p_T is contained within the subjet, leading to an overall support factor of $\delta(1-z_{\rm sub})$:

$$\begin{aligned}\Pi_A\left(z_{\rm sub};r,m_Q,p_T\right)=\frac{\alpha_s}{2\pi}\frac{\left(\mu^2e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1dz\int_0^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp<z(1-z)p_Tr\right)\\ \times\,\delta(1-z_{\rm sub})\,.\end{aligned} \tag{4.7}$$

In the above equation, ϵ is that from dimensional regularization over $d=4-2\epsilon$. The full expression for this contribution, as well as those to follow, is collected in the Appendix.

It is worth noting that the sum of (LO), (V), and (A) only has $z_{\rm sub}$-support at unity with its coefficient being the exclusive heavy-quark jet function for a jet of transverse momentum p_T and radius r

$$\delta\left(1-z_{\rm sub}\right)+\Pi_V\left(z_{\rm sub};m_Q\right)+\Pi_A\left(z_{\rm sub};r,m_Q,p_T\right)=\delta(1-z_{\rm sub})\mathcal{J}_Q(r,m_Q,p_T)\,. \tag{4.8}$$

We will make use of this relation in deriving an important sum rule obeyed by the heavy-quark subjet function.

The (B) contribution marks the situation where the splitting is resolved and the heavy quark carries the bulk of the transverse momentum. This of course means that heavy-quark trajectory defines the WTA axis and the subjet derives its $z_{\rm sub}$-support directly from the heavy-quark splitting fraction z:

$$\Pi_B\left(z_{\rm sub};r,R,m_Q,p_T\right)=\frac{\alpha_s}{2\pi}\frac{\left(\mu^2e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1dz\int_0^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp>z(1-z)p_Tr\right)$$
$$\times\,\Theta\left(\ell_\perp<z(1-z)p_TR\right)\Theta(z>1/2)\delta(z-z_{\rm sub})\,. \tag{4.9}$$

Lastly, the (C) contribution occurs when the splitting is resolved but the gluon has a greater transverse momentum fraction than the heavy quark. Therefore, the gluon forms the subjet with the WTA axis lying along its direction of propagation:

$$\Pi_C\left(z_{\rm sub};r,R,m_Q,p_T\right)=\frac{\alpha_s}{2\pi}\frac{\left(\mu^2e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1dz\int_0^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\hat{P}_{gQ\leftarrow Q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp>z(1-z)p_Tr\right)$$
$$\times\,\Theta\left(\ell_\perp<z(1-z)p_TR\right)\Theta(z>1/2)\delta(z-z_{\rm sub})\,. \tag{4.10}$$

Two constraints that follow trivially from the definitions of (B) and (C)—but are still worth drawing one's attention to—are that Π_B and $\Pi_C\,\to\,0$ as $r\,\to\,R$. These will provide inutition for the behavior of the observables in the following section.

At this stage, we have all the ingredients required to construct the heavy-quark-initiated WTA subjet function. To do so, we simply add the (LO), (V), (A), (B), and (C) phase space configurations together:

$$\mathcal{G}_Q^{\rm jet}\left(z_{\rm sub};r,R,m_Q,p_T\right)=\delta\left(1-z_{\rm sub}\right)\mathcal{J}_Q(r,m_Q,p_T)+\Pi_B\left(z_{\rm sub};r,R,m_Q,p_T\right)$$
$$+\,\Pi_C\left(z_{\rm sub};r,R,m_Q,p_T\right)\,, \tag{4.11}$$

where we have made use of Eq. (4.8). Eq. (4.11) leads directly to the important sum rule

$$\int_0^1dz_{\rm sub}\;\mathcal{G}_Q^{\rm jet}(z_{\rm sub};r,R,m_Q,p_T)=\mathcal{J}_Q(R,m_Q,p_T)\,, \tag{4.12}$$

which carries the physical meaning that if one is to integrate over all possible transverse momentum fractions that a WTA subjet of radius r can take on within a jet of radius R, one obtains

precisely the number density of jets with radius R. The proof of Eq. (4.12) can be straightforwardly demonstrated at the level of the integrands. Though elementary, the demonstration is somewhat notationally cumbersome, hence we collect it in the Appendix.

Since $\mathcal{G}_Q^{\rm jet}(z_{\rm sub}; r, R, m_Q, p_T)$ is the number density of subjets of momentum fraction $z_{\rm sub}$ with radius r about the WTA axis produced inside a jet of radius R, we can interpret Eq. (4.12) as the total number of subjets—a number which can be converted to a probability density according to

$$p\left(z_{\rm sub}; r, R, m_Q, p_T\right) \equiv \frac{\mathcal{G}_Q\left(z_{\rm sub}; r, R, m_Q, p_T\right)}{\mathcal{J}_Q\left(R, m_Q, p_T\right)}\,, \tag{4.13}$$

where Eq. (4.12) tells us immediately that

$$\int_0^1 dz_{\rm sub}\, p\left(z_{\rm sub}; r, R, m_Q, p_T\right) = 1\,. \tag{4.14}$$

Thus $p\left(z_{\rm sub}; r, R, m_Q, p_T\right)$ defines a probability distribution over the variable $z_{\rm sub}$. There exists an interesting and simplifying consequence to Eq. (4.13). Let us imagine isolating the delta-function support in the first line of Eq. (4.8) by defining coefficient functions

$$\begin{aligned}\Pi_V\left(z_{\rm sub}; m_Q\right) &\equiv \delta(1-z_{\rm sub})\,\pi_V\left(m_Q\right)\,,\\ \Pi_A\left(z_{\rm sub}; r, R, m_Q, p_T\right) &\equiv \delta(1-z_{\rm sub})\,\pi_A\left(r, m_Q, p_T\right)\,,\end{aligned} \tag{4.15}$$

so that we may express $\mathcal{J}_Q(r, m_Q, p_T)$ as

$$\mathcal{J}_Q(r, m_Q, p_T) = 1 + \pi_V\left(m_Q\right) + \pi_A\left(r, m_Q, p_T\right)\,, \tag{4.16}$$

where it is important to keep in mind that all the Π_k functions are $\mathcal{O}\left(\alpha_s\right)$, and this counting is carried in the coefficient functions π_k. Now, consider any function $f(z_{\rm sub})$ subject to the constraint that $f(1) = 1$. The average of such a function, through $\mathcal{O}\left(\alpha_s\right)$, is then

$$\begin{aligned}\langle f\rangle &= \int_0^1 dz_{\rm sub}\, f(z_{\rm sub})\, p\left(z_{\rm sub}; r, R, m_Q, p_T\right)\\ &= f(1)\Big[1 + \pi_V\left(m_Q\right) + \pi_A\left(r, m_Q, p_T\right)\Big] - \pi_V\left(m_Q\right) - \pi_A\left(R, m_Q, p_T\right)\\ &\quad + \int_0^1 dz_{\rm sub}\, f(z_{\rm sub})\Big[\Pi_B\left(z_{\rm sub}; r, R, m_Q, p_T\right) + \Pi_C\left(z_{\rm sub}; r, R, m_Q, p_T\right)\Big]\\ &= 1 + \pi_A\left(r, m_Q, p_T\right) - \pi_A\left(R, m_Q, p_T\right)\\ &\quad + \int_0^1 dz_r\, f(z_{\rm sub})\Big[\Pi_B\left(z_{\rm sub}; r, R, m_Q, p_T\right) + \Pi_C\left(z_{\rm sub}; r, R, m_Q, p_T\right)\Big]\,,\end{aligned} \tag{4.17}$$

where we made use of Eq. (4.11). What is noteworthy about Eq. (4.17) is that the (V) contribution has cancelled out and this has the effect of removing any dependence on the heavy-quark mass *itself*. The only way m_Q appears in Eq. (4.17) is on equal footing with either the scales $p_T r$ or $p_T R$. As such, all appearances of p_T can be stripped off and divide each instance of m_Q. This then results in $\langle f \rangle$ having m_Q dependence purely[1] through the "dead-cone" angle:

$$\theta_{\rm dc} \equiv \frac{m_Q}{p_T} \,. \tag{4.18}$$

Consider the generic $N^{\rm th}$ moment of the $z_{\rm sub}$ distribution, i.e.

$$\left\langle z_{\rm sub}^N \right\rangle \equiv \int_0^1 dz_{\rm sub}\, z_{\rm sub}^N\, p\left(z_{\rm sub}; r, R, m_Q, p_T\right) \,. \tag{4.19}$$

Now, $z_{\rm sub}^N$ is a function $f(z_{\rm sub})$ such that $f(1) = 1$ and therefore, by the previous argument, we see that the $N^{\rm th}$ moment depends on m_Q only through $\theta_{\rm dc}$:

$$\left\langle z_{\rm sub}^N \right\rangle = \mathcal{F}\left(N, r, R, \theta_{\rm dc}, p_T\right) \,. \tag{4.20}$$

Since any probability distribution can be expressed in a moment expansion about its mean (or first moment), this implies that the distribution $p\left(z_r; r, R, m_Q, p_T\right)$ depends on m_Q only through $\theta_{\rm dc}$:

$$p\left(z_{\rm sub}; r, R, m_Q, p_T\right) = p\left(z_{\rm sub}; r, R, \theta_{\rm dc}, p_T\right) \,. \tag{4.21}$$

This is to say that instead of being parameterized by two angles and two scales, the probability distribution over $z_{\rm sub}$ is parameterized by three angles and one scale.[2] Thus, the net effect of the heavy-quark mass is the emergence of a new angle, the dead-cone angle.

4.2.3 Shape observables and RG evolution

Armed with the heavy-quark subjet distribution defined in Eq. (4.13), we can analytically compute its moments and make direct contact with various shape observables that are of experimental interest.

[1] We should note that on dimensional grounds, one can trivially replace all instances of m_Q by $p_T \theta_{\rm dc}$, resulting in an effective "jet scale" for a quasi-jet of radius $\theta_{\rm dc}$. Such scales naturally emerge if one formally considers the $r \to 0$ limit of phase space and matches onto boosted Heavy Quark Effective Theory (bHQET), as in [MV18]. The cancellation of the virtual contributions allows for any instances of this trivial replacement by $p_T \theta_{\rm dc}$ to reduce simply to $\theta_{\rm dc}$.

[2] See previous footnote.

The simplest observable is the jet shape, which is the first cumulant, or the mean, of our $z_{\rm sub}$-distribution. Thus setting $N = 1$ in Eq. (4.19), we have

$$\Psi_Q\left(r; R, \theta_{\rm dc}\right) = \left\langle z_{\rm sub} \right\rangle , \tag{4.22}$$

where we separate the dependence on r from that on R and $\theta_{\rm dc}$ with a semi-colon to denote the fact that for any given jet, R and $\theta_{\rm dc}$ are treated as fixed, whereas r is the angular resolution variable whose variation we control to reveal the dead-cone angle.

The next cumulant to investigate is the second, or the variance, which is a non-linear combination of the first two moments. This observable, which we will refer to as the jet dispersion, can be computed as

$$\Xi_Q\left(r; R, \theta_{\rm dc}\right) = \left\langle z_{\rm sub}^2 \right\rangle - \left\langle z_{\rm sub} \right\rangle^2 . \tag{4.23}$$

We will refer to this observable as the jet dispersion. We see that while Ξ_Q probes the second moment of $p\left(z_{\rm sub}; r, R, \theta_{\rm dc}, p_T\right)$, it also has a non-linear relationship with the jet shape Ψ_Q.

Now, the heavy-quark mass m_Q constitutes an IR scale and therefore does not affect the UV evolution of the subjet function. Hence, the moments of the subjet function evolve according to the same modified DGLAP evolution as in [NSW17, KRW17, NPW19b]. As we are restricting to the case of heavy-quark-initiated jets, the only elements we need consider in our DGLAP matrix are those pertaining to splittings of light quarks

$$\begin{aligned} P_{qg\leftarrow q}(z) &= \frac{1+z^2}{(1-z)_+} + \frac{3}{2}\,\delta(1-z)\,, \\ P_{gq\leftarrow q}(z) &= \frac{1+(1-z)^2}{z}\,. \end{aligned} \tag{4.24}$$

Then at LL accuracy, we need only consider the evolution of the $N^{\rm th}$ moments of the following sum of (WTA-modified) splitting kernels

$$\Gamma(N) \equiv \int_0^1 dz_{\rm sub}\, z_{\rm sub}^N \Theta\left(z_{\rm sub} > 1/2\right) \left[P_{qg\leftarrow q}\left(z_{\rm sub}\right) + P_{gq\leftarrow q}\left(z_{\rm sub}\right) \right] . \tag{4.25}$$

For the aforementioned shape observables, we will only need the first few moments

$$\begin{aligned}\Gamma(0) &= 0\,,\\ \Gamma(1) &= C_F\left(\frac{3}{8}-2\log 2\right)\,,\\ \Gamma(2) &= C_F\left(-\frac{3}{8}-2\log 2\right)\,,\end{aligned} \tag{4.26}$$

where the first equality follows from unitarity. The natural range of scales through which we evolve our subjet functions can be obtained by examining the logarithmic form of $\mathcal{J}_Q(r, m_Q, p_T)$, which has the natural scale $\mu = \sqrt{m_Q^2 + p_T^2 r^2}$ [DKL18, KRS21, DKL21]. Thus, we evolve of the moments of $p\left(z_{\rm sub}; r, R, \theta_{\rm dc}, p_T\right)$ from $\mu_i = \sqrt{m_Q^2 + p_T^2 r^2}$ to $\mu_f = \sqrt{m_Q^2 + p_T^2 R^2}$. In terms of the dead-cone angle, we can write the high scale as $p_T R \times \sqrt{1+\theta_{\rm dc}^2/R^2}$ and the low scale as $p_T r \times \sqrt{1+\theta_{\rm dc}^2/r^2}$. Thus, $N^{\rm th}$ moment at LO+LL accuracy is given by

$$\mathcal{U}\left(r; R, \theta_{\rm dc}, p_T, N\right) \equiv \left(\frac{\alpha_s\left(p_T R \times \sqrt{1+\theta_{\rm dc}^2/R^2}\right)}{\alpha_s\left(p_T r \times \sqrt{1+\theta_{\rm dc}^2/r^2}\right)}\right)^{-2\,\Gamma(N)/\beta_0}\,, \tag{4.27}$$

where $\beta_0 = 11C_A/3 - 4n_f T_F/3$. Substituting the LO expression for α_s we then have

$$\mathcal{U}\left(r; R, \theta_{\rm dc}, N\right) \equiv \left(1 - \frac{\alpha_s\beta_0}{4\pi}\log\frac{R^2+\theta_{\rm dc}^2}{r^2+\theta_{\rm dc}^2}\right)^{-2\,\Gamma(N)/\beta_0}\,, \tag{4.28}$$

where we take α_s to be evaluated at the (constant) high scale $p_T R \times \sqrt{1+\theta_{\rm dc}^2/R^2}$. Eq. (4.28) reveals that an important effect of the dead-cone angle is to provide a shielding for the $r \to 0$ collinear singularity that is manifest in massless WTA shape observables.

At NLO, the $N^{\rm th}$ moment in Eq. (4.20) takes the form

$$\left\langle z_{\rm sub}^N \right\rangle = 1 + \frac{\alpha_s}{2\pi}\left[\Gamma(N)\log\frac{R^2+\theta_{\rm dc}^2}{r^2+\theta_{\rm dc}^2} + \mathcal{R}\left(r; R, \theta_{\rm dc}, N\right)\right] \tag{4.29}$$

where we have introduced the remainder function $\mathcal{R}\left(r; R, \theta_{\rm dc}, N\right)$, which essentially acts to contain all the power-corrections induced by the heavy-quark mass, and vanishes in the massless limit, i.e. the limit of vanishing dead-cone angle

$$\lim_{\theta_{\rm dc}\to 0}\mathcal{R}\left(r; R, \theta_{\rm dc}, N\right) = 0\,. \tag{4.30}$$

Thus, using an additive renormalization scheme, we can express the evolved moments to NLO+LL accuracy as

$$\left\langle z_{\rm sub}^{N}\right\rangle = \mathcal{U}\left(r;R,\theta_{\rm dc},N\right) + \frac{\alpha_s}{2\pi}\mathcal{R}\left(r;R,\theta_{\rm dc},N\right)\ . \tag{4.31}$$

Thus, the final integrated jet shape takes the form

$$\Psi_Q\left(r;R,\theta_{\rm dc}\right) = \mathcal{U}\left(r;R,\theta_{\rm dc},1\right) + \frac{\alpha_s}{2\pi}\mathcal{R}\left(r;R,\theta_{\rm dc},1\right)\ , \tag{4.32}$$

while the integrated momentum dispersion is given by

$$\Xi_Q\left(r;R,\theta_{\rm dc}\right) = \mathcal{U}\left(r;R,\theta_{\rm dc},2\right) + \frac{\alpha_s}{2\pi}\mathcal{R}\left(r;R,\theta_{\rm dc},2\right) - \left[\mathcal{U}\left(r;R,\theta_{\rm dc},1\right) + \frac{\alpha_s}{2\pi}\mathcal{R}\left(r;R,\theta_{\rm dc},1\right)\right]^2\ . \tag{4.33}$$

Given the evolved forms of the integrated jet shape and dispersion, we see that their differential counterparts are then given by

$$\psi_Q\left(r;R,\theta_{\rm dc}\right) = \frac{d}{dr}\,\Psi_Q\left(r;R,\theta_{\rm dc}\right)\ , \tag{4.34}$$

and

$$\xi_Q\left(r;R,\theta_{\rm dc}\right) = \frac{d}{dr}\,\Xi_Q\left(r;R,\theta_{\rm dc}\right)\ , \tag{4.35}$$

respectively. Eqs. (4.32-4.35) are the main results of this chapter.

We emphasize that the four expressions of Eqs. (4.32-4.35) are all to be thought of simply as functions three angles; one that is to be varied and two that are fixed for any given jet. The primary, in whose variation we are interested, being the angular resolution variable r, which equivalently defines the subjet radius and angular deviation about the WTA axis. The first fixed angle is just the overall jet radius R, within which the WTA subjets are to be clustered, while the second fixed angle, the dead-cone angle, captures the net effect of the heavy-quark mass. In this way, all massless analogues to Eqs. (4.32-4.35) can be obtained by taking $\theta_{\rm dc} \to 0$. This concludes the formalism portion of this chapter.

4.3 Phenomenology

In this section, we present our phenomenological predictions for the four observables, which are given to NLO+LL accuracy by the expressions of Eqs. (4.32-4.35). Our simple formalism readily

applies to sPHENIX kinematics. For pp collisions at $\sqrt{s} = 200$ GeV, one can expect jets of $p_T^{\rm jet}$ at most a few times $\sim \mathcal{O}(10)\,\mathrm{GeV}$. In this region of phase space, the non-trival power corrections contained in the remainder functions $\mathcal{R}\left(r; R, \theta_{\rm dc}, N\right)$ are non-negligible and for jets of radius $R \approx 1$ the effects of QCD evolution are minimal.[3]

Predictions for the integrated observables, $\Psi(r)$ and $\Xi(r)$, are presented in Fig. 4.2. Here, we consider jets of radius $R = 1$ and transverse momentum $p_T^{\rm jet} = 30$ GeV initiated by bottom, charm, and light quarks.

Before presenting our results, we first define the jet shape and dispersion from an experimental perspective in order to make contact with our theoretical relations. The $N^{\rm th}$ moment of the subjet-energy distribution as a cumulative distribution in r about the WTA axis is defined as

$$\mathcal{M}(N, r) \equiv \sum_{i \in \rm jet} z_i^N \Theta(r - r_{{\rm wta},i})\,. \tag{4.36}$$

Thus, we sum the $N^{\rm th}$ powers of the energy fractions z_i of the particles contained within the disk of radius r centered about the WTA axis. Thus, the differential version of the $N^{\rm th}$ moment takes the from

$$\frac{d}{dr}\mathcal{M}(N, r) \equiv \sum_{i \in \rm jet} z_i^N \delta(r - r_{{\rm wta},i})\,. \tag{4.37}$$

The interpretation for the differential observables is then that for a particle of distance r from the WTA axis, the z_i^N is placed in the bin in which r is contained. With such definitions, the integrated jet shape and dispersion are defined as

$$\begin{aligned}
\Psi(r; R, \theta_{\rm dc}) &= \mathcal{M}(1, r)\,, \\
\Xi(r; R, \theta_{\rm dc}) &= \mathcal{M}(2, r) - \mathcal{M}(1, r)^2\,,
\end{aligned} \tag{4.38}$$

while the differential versions are

$$\begin{aligned}
\Psi(r; R, \theta_{\rm dc}) &= \frac{d}{dr}\mathcal{M}(1, r)\,, \\
\Xi(r; R, \theta_{\rm dc}) &= \frac{d}{dr}\left(\mathcal{M}(2, r) - \mathcal{M}(1, r)^2\right)\,.
\end{aligned} \tag{4.39}$$

[3]This point keeps us well outside the region of phase space where $\theta_{\rm dc} << 1$, and effectively prevents any large logarithms of the form $\log\theta_{\rm dc}$ from appearing. Theoretical control of such logarithms requires further factorization and matching onto boosted Heavy Quark Effective Theory (bHQET).

By measuring these moment observables on each jet and considering an ensemble of such values, we connect back to the theoretical moment definitions of Eqs. (4.32-4.35).

Let us examine the behavior of the integrated jet shape and dispersion, as displayed in Fig. 4.2. Looking to the jet shape in the left panel, we see clearly that the effect of the heavy-quark mass is twofold. First, by examining the behavior of each curve as $r \to 0$, we see that the parton defining the WTA axis on average carries a larger fraction of the jet's energy, which is by design. More importantly though, we see that this average energy increases with the mass of the quark, as well can see by going from the light curve, and then to the charm and bottom curves. Thus, for these heavy quark curves, we can infer this parton to be the heavy quark itself, due to the limited phase space left over by the on-shell constraint. This inference is supported by the second clear feature, which is the suppressed accumulation of energy below the dead-cone radius $\theta_{\rm dc}$. This can be seen most strikingly in the curves for bottom-quark-initiated jets, where the profiles are essentially flat up to $r \sim \theta_{\rm dc}$. We see approximately similar features in the jet dispersions depicted in the right panel of Fig. 4.2, where here, the ordering of the charm and light quark curves become flipped due to the $r \to 0$ IR singularity of the light case, resulting in numerical instability. For the case of the bottom curve, we see that the larger mass results in a smaller variance in the energy-fraction of the WTA subjet as $r \to 0$. Pairing this with the behavior of the jet shape under the same limit, we conclude the the high the mass of the initiating quark, the higher fraction of the subjet energy it takes on average, and therefore, by necessity due to the finite energy budget, the lower the variance of that distribution must be. In a fashion analogous to the jet shape, as $r \to \theta_{\rm dc}$, the decrease in the dispersion is most suppressed for the heavy b-quarks. Thus, the $z_{\rm zub}$ distribution has an approximately constant variance up until the dead-cone angle, as the suppression of gluon radiation keeps the distribution of energy within this domain relatively unchanged.

Note that the integrated jet shape and dispersion are normalized such that they, respectively, approach unity and vanish as $r \to R$ across all quark flavors, but differ in their $r \to 0$ limits. This means that the rates of accumulation, or r-derivatives, must necessarily differ across flavors. This is to say that the overall normalization for the differential observables is highly flavor-dependent, a feature which can be amply seen in both panels of Fig. 4.3. Due to this fact, we also display the differential observables for charm and bottom quarks with the bottom curves scaled up to match

the normalization of the charm in Fig. 4.4.

Fig. 4.4 makes it clear that the extrema of the differential observables occur at values of $r \sim \mathcal{O}(1) \times \theta_{\rm dc}$. For the extrema to occur here is for the dead-cone angle to mark the onset of the most rapid accumulation of energy density as the subjet radius about the heavy quarks increases—a direct manifestation of the dead-cone effect. This also implies that the remaining energy of the jet, i.e. that which does not belong to the subjet, is largely localized to $r \sim \mathcal{O}(1) \times \theta_{\rm dc}$ for heavy-quark-initiated jets. This is in contrast to light-quark-initiated jets, where the subjet energy distribution exhibits power-law dependence of the form $\sim 1/r$. Such behavior is due to the approximately scale-free behavior of QCD with massless quarks. Once we analyze the angular distributions about heavy quarks, the heavy quark mass m_Q introduces an intrinsic scale, which then manifests itself through the characteristic angle $\theta_{\rm dc}$.

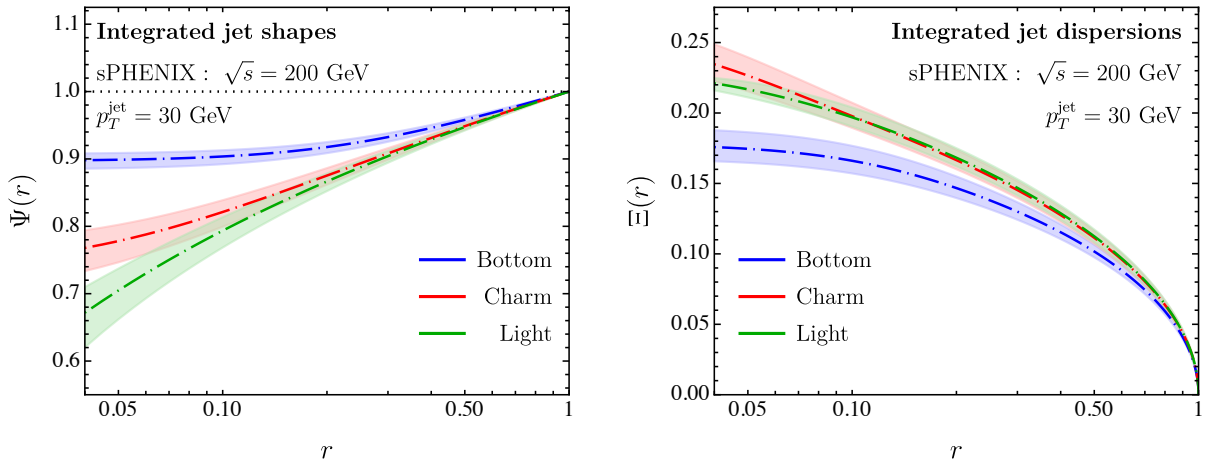

Figure 4.2: Integrated cumulants for bottom-, charm-, and light-quark-initiated jets. (Left) The first cumulant/jet shape. (Right) The second cumulant/jet dispersion.

4.4 Conclusion

In this chapter, we have proposed a pair of simple observables that reveal the dead-cone effect in a clear way. These observables are the first two cumulants of the $z_{\rm sub}$ distribution, where $z_{\rm sub}$ is the ratio of the energy of the subjet centered about the WTA axis relative to the overall jet in which it lives. These cumulants are treated as functions of the subjet radius and we consider both the

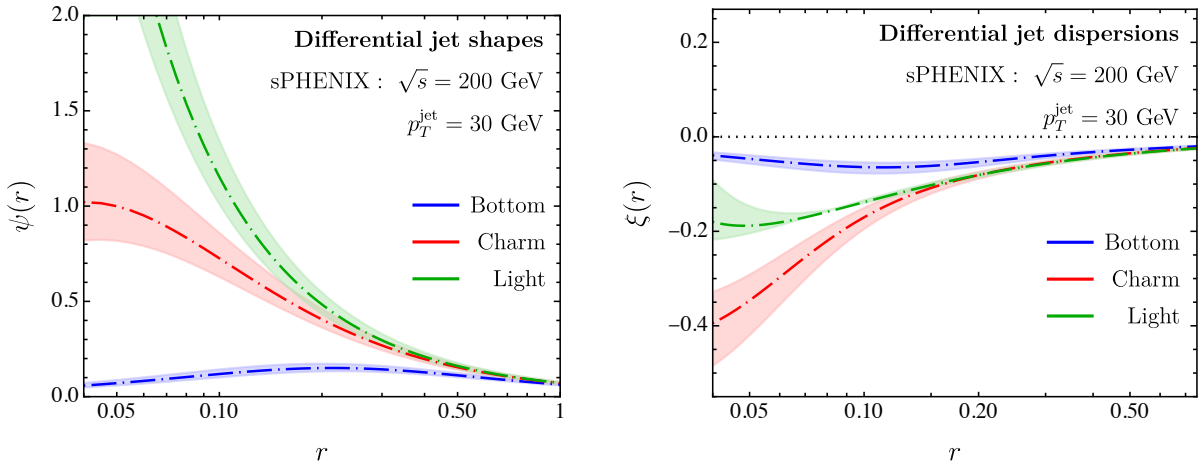

Figure 4.3: Differential cumulants for bottom-, charm-, and light-quark-initiated jets. (Left) The first cumulant/jet shape. (Right) The second cumulant/jet dispersion.

r-dependence of them as well as that of their respective r-derivatives. Those for the first cumulant correspond to what are traditionally known as the integrated and differential jet shapes, while those for the second cumulant we refer to as the integrated and differential jet dispersions. We find that the differential observables manifest the dead-cone effect in the most unambiguous way through the emergence of extrema at subjet radii of some order-one number times the dead-cone angle $\theta_{\rm dc}$. We reiterate that by enforcing the subjet to be centered about the WTA axis, we are able to preferentially pick out the trajectory of the heavy-quark initiating the jet and consider angular disks about that trajectory. We hope that the predictions laid forth in this work provide a simple and direct probe of the dead-cone effect that can be readily measured by the sPHENIX collaboration at RHIC.

Appendix

In this appendix, we first list the analytic expressions for phase space contributions to the subjet function, computed in Sec. 4.2.2. Second, we include analytic expressions the $z_{\rm sub}$-integrals that must be evaluated in obtaining the first and second moments. Lastly, we provide an explicit proof of the sum rule stated in Eq. (4.12).

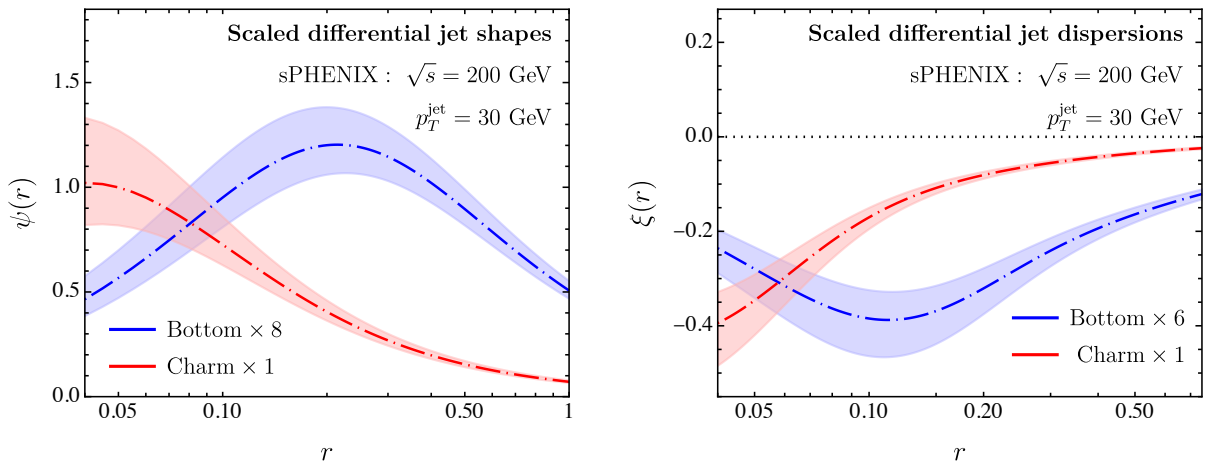

Figure 4.4: The differential observables for bottom- and charm-initiated jets. Curves for the bottom quarks are scaled so as to normalize to the charm curves and thus make the locations of their peaks at their respective values of $r \sim \theta_{\text{dc}}$ manifest. (Left) The first cumulant/jet shape. (Right) The second cumulant/jet dispersion.

The analytic expressions to the (A), (B), and (C) phase space configurations of Sec. 4.2.2 are

$$\begin{aligned}\Pi_A\left(z_{\text{sub}}; m_Q, p_T, r\right) = \delta\left(1 - z_{\text{sub}}\right) \Bigg[& \frac{1}{\epsilon}\left(\frac{p_T^2 r^2}{p_T^2 r^2 + m_Q^2} + \log\frac{m_Q^2}{p_T^2 r^2 + m_Q^2}\right) + \frac{p_T^2 r^2}{p_T^2 r^2 + m_Q^2}\log\frac{\mu^2}{p_T^2 r^2} \\ & - \frac{1}{2}\log\frac{m_Q^2}{p_T^2 + m_Q^2} + \log\frac{m_Q^2}{p_T^2 + m_Q^2}\log\frac{\mu^2}{p_T^2 r^2} - \frac{1}{2}\log^2\frac{m_Q^2}{p_T^2 r^2} - \text{Li}_2\left(\frac{m_Q^2}{p_T^2 r^2}\right) \\ & - \frac{2m_Q}{p_T r}\frac{p_T^2 r^2 + 2m_Q^2}{p_T^2 r^2 + m_Q^2}\cot^{-1}\left(\frac{m_Q}{p_T r}\right) + \frac{m_Q^2}{2p_T^2 r^2}\frac{m_Q^2 - p_T^2 r^2}{p_T^2 r^2 + m_Q^2}\log\frac{m_Q^2}{p_T^2 r^2 + m_Q^2} \\ & - 2\left\{\text{Li}_2\left(\frac{p_T r}{p_T r + i m_Q}\right) + \text{Li}_2\left(\frac{p_T r}{p_T r - i m_Q}\right)\right\} - \frac{\pi^2}{6} + \frac{9}{2}\Bigg], \end{aligned} \tag{4.40}$$

$$\begin{aligned}\Pi_B\left(z_{\rm sub};m_Q,p_T,r,R\right)=\delta(1-z_{\rm sub})\Bigg[&\frac{1}{\epsilon}\left(\frac{p_T^2R^2}{p_T^2R^2+m_Q^2}-\frac{p_T^2r^2}{p_T^2r^2+m_Q^2}-\log\frac{p_T^2R^2+m_Q^2}{p_T^2r^2+m_Q^2}\right)\\&+\log\frac{p_T^2R^2+m_Q^2}{p_T^2r^2+m_Q^2}-\frac{1}{2}\log^2\frac{R^2}{r^2}-\log\frac{R^2}{r^2}\log\frac{p_T^2r^2}{p_T^2r^2+m_Q^2}\\&-\log\frac{\mu^2}{p_T^2R^2}\log\frac{p_T^2R^2+m_Q^2}{p_T^2r^2+m_Q^2}-\frac{p_T^2r^2}{p_T^2r^2+m_Q^2}\log\frac{\mu^2}{p_T^2r^2}\\&+\frac{p_T^2R^2}{p_T^2R^2+m_Q^2}\log\frac{\mu^2}{p_T^2R^2}+\mathrm{Li}_2\left(-\frac{m_Q^2}{p_T^2r^2}\right)-\mathrm{Li}_2\left(-\frac{m_Q^2}{p_T^2R^2}\right)\Bigg]\\&+\Theta\left(z_{\rm sub}>1/2\right)\Bigg[\frac{1+z_{\rm sub}^2}{(1-z_{\rm sub})_+}\log\frac{z_{\rm sub}^2p_T^2R^2+m_Q^2}{z_{\rm sub}^2p_T^2r^2+m_Q^2}\\&+\frac{1}{(1-z_{\rm sub})_+}\frac{2m_Q^2\left(p_T^2r^2-p_T^2R^2\right)z_{\rm sub}^3}{\left(z_{\rm sub}^2p_T^2r^2+m_Q^2\right)\left(z_{\rm sub}^2p_T^2R^2+m_Q^2\right)}\Bigg],\end{aligned}\tag{4.41}$$

and

$$\begin{aligned}\Pi_C\left(z_{\rm sub};m_Q,r,R\right)=\Theta\left(z_{\rm sub}>1/2\right)\Bigg[&\frac{1+(1-z_{\rm sub})^2}{z_{\rm sub}}\log\frac{(1-z_{\rm sub})^2p_T^2R^2+m_Q^2}{(1-z_{\rm sub})^2p_T^2r^2+m_Q^2}\\&+\frac{2m_Q^2\left(p_T^2r^2-p_T^2R^2\right)(1-z_{\rm sub})^3}{\left((1-z_{\rm sub})^2p_T^2r^2+m_Q^2\right)\left((1-z_{\rm sub})^2p_T^2R^2+m_Q^2\right)}\Bigg].\end{aligned}\tag{4.42}$$

Note that in all these expressions, factors of m_Q appear on equal footing to factors of p_Tr and p_TR. This is what allows for the subjet-energy distribution to be treated as a function of the three angles r, R, and $\theta_{\rm dc}$.

The computation of the first two moments then requires the evaluation of integrals of the above expressions. These can be cast into the general for depending on a dimensionless parameter λ, which, practically speaking, takes on values of $m_Q^2/(p_Tr)^2$ and $m_Q^2/(p_TR)^2$, where the ultimate expression for each moment consists of an alternating sum between these two values, as is required for the maintenance of the $r\to R$ limit. This is automatically taken care of in some of the expressions where both λ and Λ appear—in such cases, we take $\lambda=m_Q^2/(p_Tr)^2$ and $\Lambda=m_Q^2/(p_TR)^2$. The required integral expressions are then

$$\begin{aligned}\mathcal{I}_{B1}(\lambda,1)&=\int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\frac{z_{\rm sub}}{(1-z_{\rm sub})_+}\frac{2z_{\rm sub}^3}{\lambda+z_{\rm sub}^2}\\&=-\frac{2\lambda^{3/2}\left(\cot^{-1}\left(2\sqrt{\lambda}\right)-\cot^{-1}\left(\sqrt{\lambda}\right)\right)}{\lambda+1}+\frac{\lambda^2(\log(\lambda+1)-\log(4\lambda+1))}{\lambda+1}\\&\quad+2(\lambda-1)\log(2)-\frac{7}{4},\end{aligned}\tag{4.43}$$

$$\mathcal{I}_{B1}(\lambda,2) = \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\frac{z_{\rm sub}^2}{(1-z_{\rm sub})_+}\frac{2z_{\rm sub}^3}{\lambda+z_{\rm sub}^2}$$
$$= \frac{2\lambda^{5/2}\left(\cot^{-1}\left(2\sqrt{\lambda}\right)-\cot^{-1}\left(\sqrt{\lambda}\right)\right)}{\lambda+1} + \frac{\lambda^2(\log(\lambda+1)-\log(4\lambda+1))}{\lambda+1}$$
$$+\lambda+2(\lambda-1)\log(2)-\frac{7}{3}\,, \tag{4.44}$$

$$\mathcal{I}_{B2}(\lambda,1) = \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\,z_{\rm sub}\,\frac{1+z_{\rm sub}^2}{(1-z_r)_+}\log\left(\lambda+z_{\rm sub}^2\right)$$
$$= \frac{1}{3}i\lambda^{3/2}\log\left(\frac{2\lambda-i\sqrt{\lambda}+1}{2\lambda+i\sqrt{\lambda}+1}\right) - \frac{\lambda}{3} + 2\log^2\left(1+i\sqrt{\lambda}\right) + 2\log^2\left(1-i\sqrt{\lambda}\right)$$
$$-\lambda\log(2)+\frac{1}{2}\lambda\log\left(4-\frac{3}{\lambda+1}\right)-\frac{17}{6}\log(\lambda+1)+\frac{7}{6}\log(4\lambda+1)$$
$$-2\log\left(\frac{1}{2}+i\sqrt{\lambda}\right)\log\left(2+2i\sqrt{\lambda}\right)-2\log\left(\frac{1}{2}-i\sqrt{\lambda}\right)\log\left(2-2i\sqrt{\lambda}\right)$$
$$-4\sqrt{\lambda}\tan^{-1}\left(2\sqrt{\lambda}\right)+4\sqrt{\lambda}\tan^{-1}\left(\sqrt{\lambda}\right)+2{\rm Li}_2\left(\frac{i+2\sqrt{\lambda}}{2i+2\sqrt{\lambda}}\right)+2{\rm Li}_2\left(\frac{i-2\sqrt{\lambda}}{2i-2\sqrt{\lambda}}\right)$$
$$-\frac{2\pi^2}{3}+\frac{185}{72}-\frac{7\log(2)}{3}\,, \tag{4.45}$$

$$\mathcal{I}_{B2}(\lambda,2) = \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\,z_{\rm sub}^2\,\frac{1+z_{\rm sub}^2}{(1-z_{\rm sub})_+}\log\left(\lambda+z_{\rm sub}^2\right)$$
$$= \frac{1}{3}i\lambda^{3/2}\log\left(\frac{2\lambda-i\sqrt{\lambda}+1}{2\lambda+i\sqrt{\lambda}+1}\right)+\frac{1}{2}\lambda^2\log(2)+\left(\frac{\lambda^2}{4}-\lambda-\frac{43}{12}+2\log(2)\right)\log(\lambda+1)$$
$$+\left(-\frac{\lambda^2}{4}+\lambda+\frac{251}{192}-2\log(2)\right)\log(4\lambda+1)-\frac{25\lambda}{48}-2\lambda\log(2)$$
$$-2\log\left(1+i\sqrt{\lambda}\right)\log\left(\frac{1+2i\sqrt{\lambda}}{1+i\sqrt{\lambda}}\right)-2\log\left(1-i\sqrt{\lambda}\right)\log\left(\frac{1-2i\sqrt{\lambda}}{1-i\sqrt{\lambda}}\right)$$
$$+2i\sqrt{\lambda}\log\left(\frac{2\lambda+i\sqrt{\lambda}+1}{2\lambda-i\sqrt{\lambda}+1}\right)+2{\rm Li}_2\left(\frac{i+2\sqrt{\lambda}}{2i+2\sqrt{\lambda}}\right)+2{\rm Li}_2\left(\frac{i-2\sqrt{\lambda}}{2i-2\sqrt{\lambda}}\right)$$
$$-\frac{2\pi^2}{3}+\frac{3527}{1152}+4\log^2(2)-\frac{251\log(2)}{96}\,, \tag{4.46}$$

$$\mathcal{I}_{C1}(\lambda,1) = \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\,z_{\rm sub}\,\frac{2(\Lambda-\lambda)(1-z_{\rm sub})^3}{(\Lambda+(1-z_{\rm sub})^2)\,(\lambda+(1-z_{\rm sub})^2)}$$
$$= 2\lambda\log(2)+\lambda\log(\lambda)-\lambda\log(4\lambda+1)-2\Lambda\log(2)-\Lambda\log(\Lambda)+\Lambda\log(4\Lambda+1)\,, \tag{4.47}$$

$$\begin{aligned}\mathcal{I}_{C1}(\lambda,2) &= \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\,z_{\rm sub}^2\,\frac{2(\Lambda-\lambda)(1-z_{\rm sub})^3}{(\Lambda+(1-z_{\rm sub})^2)\,(\lambda+(1-z_{\rm sub})^2)}\\ &= -2\lambda^{3/2}\cot^{-1}\left(2\sqrt{\lambda}\right)+\lambda+2\lambda\log(2)+\lambda\log(\lambda)-\lambda\log(4\lambda+1)\\ &\quad+2\Lambda^{3/2}\cot^{-1}\left(2\sqrt{\Lambda}\right)-\Lambda-2\Lambda\log(2)-\Lambda\log(\Lambda)+\Lambda\log(4\Lambda+1)\,, \end{aligned}\tag{4.48}$$

$$\begin{aligned}\mathcal{I}_{C2}(\lambda,1) &= \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\,z_{\rm sub}\,\frac{1+(1-z_{\rm sub})^2}{z_{\rm sub}}\log\left(\lambda+(1-z_{\rm sub})^2\right)\\ &= \left(2\sqrt{\lambda}-\frac{2\lambda^{3/2}}{3}\right)\cot^{-1}\left(2\sqrt{\lambda}\right)+\frac{\lambda}{3}+\frac{13}{24}\log(4\lambda+1)-\frac{37}{36}-\frac{13\log(2)}{12}\,, \end{aligned}\tag{4.49}$$

and

$$\begin{aligned}\mathcal{I}_{C2}(\lambda,2) &= \int_0^1 dz_{\rm sub}\,\Theta(z_{\rm sub}>1/2)\,z_{\rm sub}^2\,\frac{1+(1-z_{\rm sub})^2}{z_{\rm sub}}\log\left(\lambda+(1-z_{\rm sub})^2\right)\\ &= \left(2\sqrt{\lambda}-\frac{2\lambda^{3/2}}{3}\right)\cot^{-1}\left(2\sqrt{\lambda}\right)+\left(-\frac{\lambda^2}{2}+\lambda-\frac{77}{96}\right)\log(2)\\ &\quad+\left(\frac{\lambda^2}{4}-\frac{\lambda}{2}+\frac{77}{192}\right)\log(4\lambda+1)+\left(\frac{\lambda}{2}-\frac{\lambda^2}{4}\right)\log(\lambda)+\frac{13\lambda}{48}-\frac{1031}{1152}\,. \end{aligned}\tag{4.50}$$

Finally, the following is an explicit proof of the sum rule of Eq. (4.12):

$$\begin{aligned}&\int_0^1 dz_{\rm sub}\Big[\Pi_V\left(z_{\rm sub};m_Q\right)+\Pi_A\left(z_{\rm sub};r,m_Q,p_T\right)+\Pi_B\left(z_{\rm sub};r,R,m_Q,p_T\right)\\ &\qquad+\Pi_C\left(z_{\rm sub};r,R,m_Q,p_T\right)\Big]\\ &=\pi_V\left(m_Q\right)+\frac{\alpha_s}{2\pi}\frac{\left(\mu^2e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\Bigg[\int_0^{z^2(1-z)^2p_T^2r^2}\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\\ &+\int_{z^2(1-z)^2p_T^2r^2}^{z^2(1-z)^2p_T^2R^2}\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\Theta(z>1/2)\left(\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)+\hat{P}_{gQ\leftarrow Q}\left(z,\ell_\perp\right)\right)\Bigg]\,. \end{aligned}\tag{4.51}$$

Note that

$$\begin{aligned}&\Theta(z>1/2)\left(\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)+\hat{P}_{gQ\leftarrow Q}\left(z,\ell_\perp\right)\right)\\ &\quad=\Theta(z>1/2)\left(\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)+\hat{P}_{Qg\leftarrow Q}\left(1-z,\ell_\perp\right)\right)\,,\\ &\quad=\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\left(\Theta(z>1/2)+\Theta(z<1/2)\right)\,,\\ &\quad=\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\,, \end{aligned}\tag{4.52}$$

hence

$$\begin{aligned}
&\int_0^1 dz_{\rm sub}\Big[\Pi_V\left(z_{\rm sub};m_Q\right)+\Pi_A\left(z_{\rm sub};m_Q,p_T,r\right)+\Pi_B\left(z_{\rm sub};r,R,m_Q,p_T\right)\\
&+\Pi_C\left(z_{\rm sub};r,R,m_Q,p_T\right)\Big]\\
&=\pi_V\left(m_Q\right)+\frac{\alpha_s}{2\pi}\frac{\left(\mu^2e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\left[\int_0^{z^2(1-z)^2p_T^2r^2}+\int_{z^2(1-z)^2p_T^2r^2}^{z^2(1-z)^2p_T^2R^2}\right]\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\,,\\
&=\pi_V\left(m_Q\right)+\frac{\alpha_s}{2\pi}\frac{\left(\mu^2e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\int_0^{z^2(1-z)^2p_T^2R^2}\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\,,\\
&=\mathcal{J}_Q(m_Q,p_T,R)\,,
\end{aligned}\tag{4.53}$$

where we have defined

$$\pi_V\left(m_Q\right)\equiv\int_0^1 dz_{\rm sub}\,\Pi_V\left(z_{\rm sub};m_Q\right)\,.\tag{4.54}$$

CHAPTER 5

An Infrared flavor for heavy-quark jets

In this chapter, we study a particular definition for the flavor of jets containing heavy quarks along with its statistical and physical corollaries. The central object under consideration is the distribution of energy fractions of the Winner-Take-All axis relative to the energy of a given jet's progenitor. Such a distribution relies on both ultraviolet and infrared scales, where the heavy quark mass, m_Q, provides a crucial addition to the low-energy data required for specification. On the statistical front, we demonstrate how such a flavor definition in the infrared allows for the Bayesian inference of the flavor of the jet's progenitor in the deep ultraviolet—this allows us to determine the physical situations under which we can expect heavy-quark jets to be generated during the showering process versus emerging directly from the hard-scattering of a given collision. On the physical front, we define new observables based on cumulants of the aforementioned distribution, related to what are classically known as differential jet shapes, whose profiles are determined exclusively through infrared data, with all ultraviolet dependence being carried by their normalizations. This separation of dependencies leads to the observables exhibiting a high degree of universality across collisional systems and center-of-mass energies. We achieve this universality through the consideration of transverse momenta relative to the Winner-Take-All axis as opposed to any sort of angular deviation. This leads to a novel probe of the so-called "dead-cone" effect that can be equally-well observed at sPHENIX and LHC kinematics. We provide predictions for both.

5.1 Introduction

Jets are rich objects whose origin lies in the deep ultraviolet (UV) but whose observation takes place solely in the infrared (IR). One objective of the field of jet substructure can then be characterized as understanding the precise way in which information pertaining to a jet's UV genesis is impressed upon its final IR measurement, see e.g. [LMN17] for a review. One piece of information of particularly high interest is the flavor of the parton initiating a given jet—such information gives direct access to the hard-scattering process of the event from which the jet emerged. However, the subsequent radiative cascade which evolves the individual hard partons into the complex multi-particle systems that make up jets greatly obfuscates possible connection to this UV flavor.

One precise way to connect the IR flavor of a jet to the UV flavor of its initiating parton is given by [CLM22a] and whose factorization structure in perturbative Quantum Chromodynamics (QCD) is described in [LN23]. This construction is based on defining the IR flavor of a jet to be that of the net flow of flavor directed along the Winner-Take-All (WTA) axis [BCT14b, LNT14], which, in turn, is defined as the axis oriented along the trajectory of the most energetic particle contained in a jet. The theoretical object one computes to make statements about this flavor defintion is the distribution of energy fractions of the WTA axis with respect to the energy of the progenitor of the jet—which we will denote as $z_{\mathrm{w}} \equiv p_{\perp \mathrm{WTA}}/p_{\perp J}$—binned by flavor. This flavor binning can be understood at $\mathcal{O}(\alpha_s)$ as originating from phase-space integrations of the Altarelli-Parisi splitting kernels [AP77], from which the IR flavor-binning follows from the final state of the splitting kernels and the UV binning from the initial. Thus, the final object will be a joint distribution in z_{w} and f_{ir} which is conditioned on f_{uv}. The IR flavor is then paired with the IR scale, which we will denote as $k_\perp$ and take to be the scale at which the parton shower terminates but is still high enough to be faithfully computed using quark and gluon degrees of freedom in perturbative QCD. Analogously then, the UV flavor will be paired with the scale $p_{\perp J}$ defining the transverse momentum of the parton initiating the jet under consideration. For the case of massless partonic flavors, these are the only two scales affiliated with such IR/UV characterizations.

The focus of this chapter is to extend the above construction to accommodate massive partonic flavors, that is, for jets whose IR flavor can be reliably classified as that of a heavy quark, such

as a charm or bottom. What such an extension entails may be immediately intuited from the previous discussion. The primary, and perhaps defining, physical attribute which distinguishes heavy quarks from their light counterparts is their one-to-several GeV mass m_Q. As such, with the addition of heavy quarks to this instance of jet flavor comes the associated addition of the IR scale $m_Q \sim k_\perp$. For the massless cases, the absence of this additional scale makes it so that dependence on $k_\perp$ is purely logarithmic, whereas in the case of massive flavors, highly non-linear functions involving both $k_\perp$ and m_Q appear besides logarithms in the next-to-leading-order (NLO) computation. We will see that these $\mathcal{O}(\alpha_s)$ corrections imprint themselves in important ways onto observables affiliated with jets categorized by our IR heavy-flavor definition.

A general question whose answer is highly consequential for the maintenance and development of effectual observable design is: how are heavy flavor jets sourced? That is, are their heavy-flavor quarks predominantly acquired through process of QCD evolution? Or do they rather stem directly from the hard-scattering itself? Sourcing from the latter situation underpins how well a particular jet observable is able to probe the flavor-dependence of partonic energy loss in heavy-ion collisions, as the traversal of the quark-gluon plasma by hard heavy quarks is what gives rise to the modification in their subsequent radiation pattern [HKV13b, HKV15b, LV19c, LV19b, DWZ18b, KRV19b]. Similar hard-scattering considerations can be made to connect the final state to the initial state in order to probe important aspects of the proton, such as the Sivers function [KRS21].

We answer the above question in the following way. By marginalizing our joint distributions over WTA energy fraction z_{w}, we obtain conditional probabilities for IR flavors given UV ones. Through the use of Bayes theorem, we then construct the posterior probabilities for a UV flavor conditional on the observation of an IR flavor. Thus, allowing us to quantitatively infer the UV origin of a heavy-flavor jet in the IR. Given the probability for a particular UV source, we may then move on to infer the role played by QCD evolution sourcing the heavy quark.

Beyond marginalizing our distribution, we compute all its moments and make contact with familiar jet substructure observables, such as the jet shape, which is related to the first moment of our distribution. The jet shape is defined as in [Cha12] and has been computed using both traditional QCD techniques [LLY11, LLY13] and in SCET [EVW10, CV14b, CRW19b] about a central axis. That as centered about the WTA axis and for massless partons has been computed

[KRW17, NPW19a]. The massless jet shape is of interest in heavy-ion collisions [CV16a]. In the previous chapter, it is shown how the jet shape centered about the WTA axis is particularly well-suited for revealing the so-called "dead-cone" angle for jets initiated by heavy quarks. The dead-cone effect [DKT91] essentially corresponds to the emergence of a characteristic angle about a heavy quark, within which gluon radiation is greatly attenuated. This angle is typically denoted as $\theta_{\rm dc} \sim m_Q/E_Q$ where E_Q is the energy of the heavy quark under consideration. This brings us to an important point. A feature of the traditional jet shape is that it is defined with respect to an angular scale measured about a particular axis. Considering the case of a jet initiated by a heavy quark then, $p_{\perp J}$ plays the role of E_Q and the dead-cone angle is then a ratio of an IR scale to a UV scale. As such, $\theta_{\rm dc}$ has strong dependence on the collision system at hand, namely its center-of-mass energy, as this quantity markedly affects the observed values of $p_{\perp J}$, and hence the UV flavor of a jet.

What distinguishes the type of jet shape laid forth in this work, is its dependence not on an angular region about the WTA, but on momenta transverse to it. As such, the resulting dead-cone effect manifests itself not through any cone whose radius is dependent on UV details, but rather an absolute IR scale of the order of the heavy quark mass itself. In our formalism, the profile of the jet shape depends solely on the two IR scales $k_\perp$ and m_Q and thus makes no reference to the UV process which leads to the IR distribution of energy within the substructure of a given jet. All dependence on UV physics purely affects the overall normalization of the resultant jet shape through the effects of renormalization group (RG) flow from the UV to the IR. This results in an observable whose behavior exhibits a high degree of universality across kinematic setups, such as those achieved at the future sPHENIX experiment and those at the current LHC. This universal structure extends beyond the jet shape, i.e. the first moment/cumulant of our distribution, and into higher cumulants such as the variance.

This chapter is organized as follows. Computational details are provided in Sec. 5.2, with Sec. 5.2.1 containing those regarding massless flavors, where those for massive flavors are the work of Sec. 5.2.2. We go through the full massless case first for the sake of completion, but more importantly to lay the groundwork for the more complicated massive calculation, as these results can be elegantly expressed in terms of their massless counterparts with the addition of cor-

rection terms in which all mass-dependence is relegated. Sec. 5.3 then describes the RG evolution equations obeyed by the marginal distributions and contains our application to the Bayesian inference for the UV flavor sourcing of IR heavy-quark jets. In Sec. 5.4, we solve the RG equations governing the the evolution of the higher moments of the flavor-binned $z_{\rm w}$-distribution and make phenomenological predictions for our new jet shape observables. We conclude in Sec. 7.5 and collect analytic expressions for the moments of each UV-to-IR flavor channel in the Appendix.

5.2 Joint distributions in $f_{\rm ir}$ and $z_{\rm w}$, and their $z_{\rm w}$-moments

In this section, we compute the joint probability distributions in the two random variables of interest, which are the infrared flavor for a given jet, $f_{\rm ir}$, and the momentum-fraction of the said jet carried along the WTA axis, $z_{\rm w}$. We furthermore calculate these joint distributions to be conditional on the various possible UV flavors for the jet, $f_{\rm uv}$, all to $\mathcal{O}(\alpha_s)$. Once obtained, we compute moments in the $z_{\rm w}$ variable for a completely general moment N. Such objects evolve according to modified DGLAP-evolution equations in Mellin space. To reiterate, the final objects are $z_{\rm w}$-moments binned by IR flavor and conditioned on UV flavor.

Of course, when we make reference to IR and UV flavors, we are implicitly referring to labels that are defined at particular energy scales. To make this fact explicit, we notationally express the flavor labels as functions of their respective scales. The IR scale is taken to be $k_\perp$, which will be the resolution scale of the $\mathcal{O}(\alpha_s)$ splitting. Ultimately, dependence on $k_\perp$ is obtained through various limits of phase-space integrations. The UV scale is chosen to be $p_{\perp J}$, which is the overall transverse momentum of the jet—equivalently the scale of the partons emerging from the hard scattering process, which is $\mathcal{O}\left(\sqrt{s}\right)$ of the collision. Dependence on $p_{\perp J}$ is acquired through the QCD evolution of the $z_{\rm w}$-moments.

For the case of massless partonic IR jet flavors, we take $k_\perp \gtrsim 1$ GeV, so that it can be treated as the final scale of the parton shower while also allowing for perturbative calculations to remain meaningful. For massive flavors, we treat $k_\perp \sim m_Q$ and thus maintain all expressions containing appearances of both. Such appearances take the form of "mass-corrections" for heavy quark flavors relative to their massless counterparts, and correspondingly vanish in the $m_Q \to 0$ limit. It is

important to note that as a heavy quark mass constitutes IR information, it should have no net effect on the sum of anomalous dimensions and ultimate renormalization-group consistency of our framework. We not only confirm this to be the case, but further find m_Q to imprint itself on none of the individual anomalous dimensions either. This makes the aforementioned mass-corrections affect only the finite $\mathcal{O}(\alpha_s)$ terms in all instances involving heavy flavors.

5.2.1 Calculation for purely massless partonic flavors

As this work is to $\mathcal{O}(\alpha_s)$, the fundamental objects under study are the LO Alterelli-Parisi splitting functions. Utilizing dimensional-regularization in $d = 4 - 2\epsilon$ spacetime dimensions, we specifically start with the ϵ-dependent ones:

$$\begin{aligned}
\hat{P}_{qg\leftarrow q}\left(z,\ell_\perp\right) &= \frac{C_F}{\ell_\perp^2}\left(\frac{1+z^2}{1-z} - \epsilon\left(1-z\right)\right) , \\
\hat{P}_{gq\leftarrow q}\left(z,\ell_\perp\right) &= \frac{C_F}{\ell_\perp^2}\left(\frac{1+(1-z)^2}{z} - \epsilon\, z\right) , \\
\hat{P}_{q\bar{q}\leftarrow g}\left(z,\ell_\perp\right) &= \frac{T_F}{\ell_\perp^2}\left(z^2+(1-z)^2 - 2\epsilon\, z(1-z)\right) , \\
\hat{P}_{gg\leftarrow g}\left(z,\ell_\perp\right) &= \frac{2C_A}{\ell_\perp^2}\left(\frac{z}{1-z} + \frac{1-z}{z} + z(1-z)\right) .
\end{aligned} \tag{5.1}$$

Here, z is the energy-fraction of the split parton j relative to the initial i, referenced in the form of the subscript $jk \leftarrow i$. The variable $\ell_\perp$ is the momentum between partons j and k, transverse to that of i. The color factors are C_F, T_F, and C_A are those of $\mathrm{SU}(3)$, with values $4/3$, $1/2$, and 3, respectively. The joint distributions in f_{ir} and z_{w}, conditioned on f_{uv}, are (up to prefactors) obtained through the integration of these splitting functions over their arguments.

Let us first consider the conditional probability for a quark in the UV to split into a quark in the IR, according the WTA constraint. Such a distribution has two contributions, those arising from resolved and unresolved splittings. For the splitting to be resolved is to say that it occurs above the scale $k_\perp$ and that the quark carries the larger energy fraction of the two daughter partons. The bare

resolved contribution is

$$
\begin{aligned}
&p_\epsilon^{(\mathrm{R})}\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp) \,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{qg\leftarrow q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\Theta(z>1/2)\delta(z-z_{\mathrm{w}})\,, \\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{qg\leftarrow q}\left(z_{\mathrm{w}},\ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\Theta(z_{\mathrm{w}}>1/2)\,,
\end{aligned}
\tag{5.2}
$$

where the superscript (R) specifies that this is the resolved contribution and the subscript ϵ is to remind us that we have not renormalized the distribution—renormalization will be carried out at the level of moments.

The unresolved contribution comes from the situation where the splitting takes place below the scale $k_\perp$, leaving the trajectory of the initiating quark from the UV unperturbed, and thus leading to the net flavor along the WTA axis in the IR to remain that of a quark. This manifests itself through the overall z_{w}-support being pinned at unity:

$$
\begin{aligned}
&p_\epsilon^{(\mathrm{U})}\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp) \,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= \delta(1-z_{\mathrm{w}})\,\frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{qg\leftarrow q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp < k_\perp\right)\,.
\end{aligned}
\tag{5.3}
$$

Another process whose z_{w}-support is defined by $\delta(1-z_{\mathrm{w}})$ at $\mathcal{O}(\alpha_s)$, but would be considered distinct from an unresolved splitting, would be the emission and re-absorption of a virtual gluon by the UV quark. However, in this case, as well as all the cases involving purely massless partons, the resulting contribution takes the form of a scaleless integral and therefore vanishes in dimensional regularization. This will cease to be the case only once we include heavy flavors in Sec. 5.2.2.

Eqs. (5.2) and (5.3) provide the $\mathcal{O}(\alpha_s)$ corrections to the leading order distribution for a UV quark to split into an IR quark, that is, through no splitting at all. Such a flavor-invariant process has the LO distribution $\delta(1-z_{\mathrm{w}})$. The full conditional probability will be the sum of this LO distribution with its resolved and unresolved corrections:

$$
\begin{aligned}
&p_\epsilon\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp) \,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= \delta(1-z_{\mathrm{w}}) + p_\epsilon^{(\mathrm{R})}\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp) \,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right) + p_\epsilon^{(\mathrm{U})}\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp) \,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right)\,.
\end{aligned}
\tag{5.4}
$$

We will consider $z_{\rm w}$-moments of this conditional probability distribution

$$p_\epsilon\left(N, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) \equiv \int_0^1 dz_{\rm w}\, z_{\rm w}^N\, p_\epsilon\left(z_{\rm w}, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\,, \tag{5.5}$$

where, because we are keeping N general, we are effectively swapping $z_{\rm w}$-dependence for N-dependence, hence the choice of notation. In what follows, we will consider specific integer instances of N and express this choice in the same argument slot. Since we will only be doing this for N and never for $z_{\rm w}$, this should not lead to confusion.

In the rest of our analysis, we will treat the cases of $N = 0$ and $N > 0$ separately, as their physical interpretations are fundamentally different. To begin, let us consider the zeroth moment

$$\begin{aligned} &p_\epsilon\left(0, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) \\ &\quad = 1 + \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz_{\rm w}\int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{qg\leftarrow q}\left(z_{\rm w}, \ell_\perp\right)\Big[\Theta\left(\ell_\perp > k_\perp\right)\Theta(z_{\rm w} > 1/2) \\ &\qquad\qquad + \Theta\left(\ell_\perp < k_\perp\right)\Big]\,, \end{aligned} \tag{5.6}$$

Now, using the facts that

$$\begin{aligned} \Theta(z_{\rm w} < 1/2) + \Theta(z_{\rm w} > 1/2) &= 1\,, \\ \Theta\left(\ell_\perp < k_\perp\right) + \Theta\left(\ell_\perp > k_\perp\right) &= 1\,, \end{aligned} \tag{5.7}$$

and that scaleless integrals vanish in dimensional regularization, Eq. (5.6) reduces to

$$p_\epsilon\left(0, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) = 1 + \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^{1/2} dz_{\rm w}\int_0^{k_\perp^2} \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{qg\leftarrow q}\left(z_{\rm w}, \ell_\perp\right)\,. \tag{5.8}$$

We can see that this expression avoids the $z_{\rm w} \to 1$ singularity of the splitting function (see the first line of Eq. (5.1)) through the $z_{\rm w}$-integration and acquires IR poles through the $\ell_\perp$-integration. Expanding in ϵ and subtracting off the divergences, we obtain the renormalized zeroth moment

$$p\left(0, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) = 1 + \frac{\alpha_s C_F}{2\pi}\left[\left(\frac{5}{8} - 2\log 2\right)\log\frac{\mu^2}{k_\perp^2} + \frac{3}{8}\right]\,, \tag{5.9}$$

whose anomalous dimension is then given by

$$\gamma\left(0, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) = \frac{\alpha_s C_F}{\pi}\left(\frac{5}{8} - 2\log 2\right)\,, \tag{5.10}$$

where we port over the same notation to highlight the fact that γ too is a function of the moment number, as well as the flavors.

Now, when it comes to the $N \geq 0$ moments for flavor-invariant processes such as this, the following calculational trick is useful. Consider taking $z_{\rm w} \to 1 - z_{\rm w}$, so that our moment calculation becomes

$$p_\epsilon \left(N, q_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) = \int_0^1 dz_{\rm w}\, (1-z_{\rm w})^N\, p_\epsilon \left(1-z_{\rm w}, q_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) \,. \tag{5.11}$$

Performing the binomial expansion, we have

$$(1-z_{\rm w})^N = 1 + \sum_{J=1}^{N} \binom{N}{J} (-1)^J z_{\rm w}^J \,, \tag{5.12}$$

so that,

$$\begin{aligned} &p_\epsilon \left(N, q_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) \\ &\quad = p_\epsilon \left(0, q_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \sum_{J=1}^{N} \binom{N}{J} (-1)^J \int_0^1 dz_{\rm w}\, z_{\rm w}^J\, p_\epsilon^{(\rm R)} \left(1-z_{\rm w}, q_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) \,, \end{aligned} \tag{5.13}$$

where we have used the fact that the $z_{\rm w}$-moments of the LO and unresolved contributions to the distribution vanish under the mapping of $z_{\rm w} \to 1 - z_{\rm w}$, as this maps $\delta(1-z_{\rm w}) \to \delta(z_{\rm w})$. Let us now examine the integral appearing in the second term on the RHS

$$\begin{aligned} &\int_0^1 dz_{\rm w}\, z_{\rm w}^J\, p_\epsilon^{(\rm R)} \left(1-z_{\rm w}, q_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) \\ &\qquad = \frac{\alpha_s}{2\pi} \frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)} \int_0^{1/2} dz_{\rm w} \int_{k_\perp^2}^{\infty} \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\, z_{\rm w}^J\, \hat{P}_{gq\leftarrow q}\left(z_{\rm w}, \ell_\perp\right) \,, \end{aligned} \tag{5.14}$$

where we have made use of the facts that $\hat{P}_{qg\leftarrow q}\left(1-z_{\rm w}, \ell_\perp\right) = \hat{P}_{gq\leftarrow q}\left(z_{\rm w}, \ell_\perp\right)$ and $\Theta(1 - z_{\rm w} > 1/2) = \Theta(z_{\rm w} < 1/2)$. Now, in Eq. (5.14) we see that since $J \geq 1$, the factor of $z_{\rm w}^J$ shields the integrand from the $z_{\rm w} \to 0$ singularity of the splitting function (see the second line of Eq. (5.1).

Thus the renormalized higher moments are[1]

$$
\begin{aligned}
&p_\epsilon\left(N, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\\
&\quad= p_\epsilon\left(0, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \frac{\alpha_s C_F}{2\pi}\sum_{J=1}^{N}\binom{N}{J}(-1)^J\left(\frac{2^{-J-2}(J(5J+17)+16)}{J(J+1)(J+2)}\log\frac{\mu^2}{k_\perp^2}\right.\\
&\qquad\qquad\left. - \frac{2^{-J-2}}{J+2}\right).
\end{aligned}
\tag{5.15}
$$

The anomalous dimension is then

$$
\begin{aligned}
&\gamma\left(N, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\\
&\quad= \gamma\left(0, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \frac{\alpha_s C_F}{\pi}\sum_{J=1}^{N}\binom{N}{J}(-1)^J\frac{2^{-J-2}(J(5J+17)+16)}{J(J+1)(J+2)}\,.
\end{aligned}
\tag{5.16}
$$

Next, let us consider the conditional probability for a quark flavor in the UV to split to a gluon flavor in the IR. Such a situation requires a resolved splitting

$$
\begin{aligned}
&p_\epsilon^{(\rm R)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\int_0^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gq\leftarrow q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\Theta(z>1/2)\,\delta(z-z_{\rm w})\,,\\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_{k_\perp^2}^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gq\leftarrow q}\left(z_{\rm w},\ell_\perp\right)\Theta(z_{\rm w}>1/2)\,,
\end{aligned}
\tag{5.17}
$$

and this is, in fact, the sole contribution to $\mathcal{O}(\alpha_s)$. Thus we have

$$
p_\epsilon\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) = p_\epsilon^{(\rm R)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\,. \tag{5.18}
$$

As before, starting with the zeroth moment

$$
p_\epsilon\left(0, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) = \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_{1/2}^1 dz_{\rm w}\int_{k_\perp^2}^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gq\leftarrow q}\left(z_{\rm w},\ell_\perp\right)\,, \tag{5.19}
$$

where the WTA constraint avoids the $z_{\rm w}\to 0$ singularity of the splitting function. Its renormalized value is then

$$
p\left(0, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) = -\frac{\alpha_s C_F}{2\pi}\left[\left(\frac{5}{8}-2\log 2\right)\log\frac{\mu^2}{k_\perp^2}+\frac{3}{8}\right]\,, \tag{5.20}
$$

[1]We note here that the sum appearing in Eq. (5.15) can, in fact, be evaluated in terms of hypergeometric functions, however the form is unenlightening and far less succinct than that presented above.

with anomalous dimension

$$\gamma\left(0, g_{\text{ir}}(k_\perp) \,|\, q_{\text{uv}}(p_{\perp J})\right) = -\frac{\alpha_s C_F}{\pi}\left(\frac{5}{8} - 2\log 2\right) . \tag{5.21}$$

Examining the pair Eqs. (5.20, 5.21) in light of the pair Eqs. (5.9, 5.10), one sees immediately that

$$\begin{aligned} 1 &= p\left(0, q_{\text{ir}}(k_\perp) \,\middle|\, q_{\text{uv}}(p_{\perp J})\right) + p\left(0, g_{\text{ir}}(k_\perp) \,\middle|\, q_{\text{uv}}(p_{\perp J})\right) , \\ 0 &= \gamma\left(0, q_{\text{ir}}(k_\perp) \,|\, q_{\text{uv}}(p_{\perp J})\right) + \gamma\left(0, g_{\text{ir}}(k_\perp) \,|\, q_{\text{uv}}(p_{\perp J})\right) , \end{aligned} \tag{5.22}$$

as expected through unitarity. In general, given a fixed UV flavor f_{uv}, unitarity requires that

$$\begin{aligned} 1 &= \sum_{f_{\text{ir}}} p\left(0, f_{\text{ir}}(k_\perp) \,\middle|\, f_{\text{uv}}(p_{\perp J})\right) , \\ 0 &= \sum_{f_{\text{ir}}} \gamma\left(0, f_{\text{ir}}(k_\perp) \,|\, f_{\text{uv}}(p_{\perp J})\right) , \end{aligned} \tag{5.23}$$

making Eq. (5.22) an important consistency check.

$$\begin{aligned} 1 &= \sum_{f_{\text{ir}}} p\left(0, f_{\text{ir}}(k_\perp) \,\middle|\, f_{\text{uv}}(p_{\perp J})\right) , \\ &= \sum_{f_{\text{ir}}} \int_0^1 dz_{\text{w}}\, p\left(z_{\text{w}}, f_{\text{ir}}(k_\perp) \,\middle|\, f_{\text{uv}}(p_{\perp J})\right) . \end{aligned} \tag{5.24}$$

In contrast to the quark-to-quark case, computing higher moments of flavor-changing distributions is completely straightforward, as the resulting integrands carry no singularities in z_{w}. Performing the calculation, the higher moments for the quark-to-gluon channel are given by

$$\begin{aligned} &p\left(N, g_{\text{ir}}(k_\perp) \,\middle|\, q_{\text{uv}}(p_{\perp J})\right) \\ &\quad = \frac{\alpha_s C_F}{2\pi}\left[\left(-\frac{2^{-N-2}\left(5N^2+17N+16\right)}{N(N+1)(N+2)} + \frac{2}{N} - \frac{2}{N+1} + \frac{1}{N+2}\right)\log\frac{\mu^2}{k_\perp^2}\right. \\ &\qquad\qquad \left. + \frac{2^{-N}-4}{4(N+2)}\right] , \end{aligned} \tag{5.25}$$

and have anomalous dimensions

$$\gamma\left(N, g_{\text{ir}}(k_\perp) \,|\, q_{\text{uv}}(p_{\perp J})\right) = \frac{\alpha_s C_F}{\pi}\left(-\frac{2^{-N-2}\left(5N^2+17N+16\right)}{N(N+1)(N+2)} + \frac{2}{N} - \frac{2}{N+1} + \frac{1}{N+2}\right) . \tag{5.26}$$

This concludes the calculations that result from a light quark in the UV.

Next, we consider the situations emanating from an initial gluon in the UV. The process of a UV gluon splitting to a resolved gluon in the IR is a unique case

$$
\begin{aligned}
&p_\epsilon^{(\mathrm{R})}\left(z_{\mathrm{w}}, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gg\leftarrow g}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\left[\Theta(z>1/2)+\Theta(z<1/2)\right] \\
&\qquad\qquad\qquad\qquad \times \delta(z-z_{\mathrm{w}})\,, \\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_{k_\perp^2}^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gg\leftarrow g}\left(z_{\mathrm{w}},\ell_\perp\right)\,,
\end{aligned}
\tag{5.27}
$$

where, the identical nature of the two daughter gluons effectively lifts the WTA constraint. In the case of unresolved splittings, there exist contributions arising from both C_A and T_F channels

$$
\begin{aligned}
&p_\epsilon^{(\mathrm{U})}\left(z_{\mathrm{w}}, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= \delta(1-z_{\mathrm{w}})\,\frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\left[\hat{P}_{gg\leftarrow g}\left(z,\ell_\perp\right)+n_f\hat{P}_{q\bar{q}\leftarrow g}\left(z,\ell_\perp\right)\right]\Theta\left(\ell_\perp < k_\perp\right) \\
&\quad= \delta(1-z_{\mathrm{w}})\,\frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^{k_\perp^2} \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\left[\hat{P}_{gg\leftarrow g}\left(z,\ell_\perp\right)+n_f\hat{P}_{q\bar{q}\leftarrow g}\left(z,\ell_\perp\right)\right]\,,
\end{aligned}
\tag{5.28}
$$

where n_f denotes the number of light quark flavors. The total conditional probability is then

$$
\begin{aligned}
&p_\epsilon\left(z_{\mathrm{w}}, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= \delta(1-z_{\mathrm{w}}) + p_\epsilon^{(\mathrm{R})}\left(z_{\mathrm{w}}, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right) + p_\epsilon^{(\mathrm{U})}\left(z_{\mathrm{w}}, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right)\,.
\end{aligned}
\tag{5.29}
$$

As always, let us consider the zeroth moment

$$
\begin{aligned}
&p_\epsilon\left(0, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right) \\
&\quad= 1 + \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz_{\mathrm{w}}\left(\int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gg\leftarrow g}\left(z_{\mathrm{w}},\ell_\perp\right) + \int_0^{k_\perp^2}\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,n_f\hat{P}_{q\bar{q}\leftarrow g}\left(z_{\mathrm{w}},\ell_\perp\right)\right)\,, \\
&\quad= 1 + \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz_{\mathrm{w}}\int_0^{k_\perp^2}\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,n_f\hat{P}_{q\bar{q}\leftarrow g}\left(z_{\mathrm{w}},\ell_\perp\right)\,.
\end{aligned}
\tag{5.30}
$$

Here, the unresolved C_A channel combines with the resolved component to give a scale-free, and therefore vanishing, $\ell_\perp$-integral. Hence, only the unresolved T_F channel contributes. In the end, the renormalized zeroth moment is

$$
p\left(0, g_{\mathrm{ir}}(k_\perp) \,\middle|\, g_{\mathrm{uv}}(p_{\perp J})\right) = 1 - \frac{\alpha_s n_f T_F}{2\pi}\left(\frac{2}{3}\log\frac{\mu^2}{k_\perp^2} - \frac{1}{3}\right)\,,
\tag{5.31}
$$

and has anomalous dimension

$$\gamma\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) = -\frac{2}{3}\frac{\alpha_s n_f T_F}{\pi}\,. \tag{5.32}$$

Now we will move onto higher moments of the gluon-to-gluon channel. For this UV-to-IR non-flavor-changing process, we will make use of the same trick as utilized in the quark-to-quark case, namely

$$\begin{aligned}
&p_\epsilon\left(N, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) \\
&\quad= \int_0^1 dz_{\rm w}\,(1-z_{\rm w})^N\, p_\epsilon\left(1-z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,, \\
&\quad= p_\epsilon\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) + \sum_{J=1}^N \binom{N}{J}(-1)^J \int_0^1 dz_{\rm w}\, z_{\rm w}^J\, p_\epsilon^{(\rm R)}\left(1-z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,,
\end{aligned} \tag{5.33}$$

which can be seen as the gluonic analogue to Eq. (5.11). Let us examine the integral appearing in the second term on the RHS:

$$\begin{aligned}
&\int_0^1 dz_{\rm w}\, z_{\rm w}^J\, p_\epsilon^{(\rm R)}\left(1-z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) \\
&\qquad= \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz_{\rm w}\int_{k_\perp^2}^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\, z_{\rm w}^J\, \hat{P}_{gg\leftarrow g}\left(z_{\rm w}, \ell_\perp\right)\,,
\end{aligned} \tag{5.34}$$

where we have made use of the fact that $\hat{P}_{gg\leftarrow g}\left(z_{\rm w}, \ell_\perp\right)$ is invariant under the exchange of $z_{\rm w}$ with $1-z_{\rm w}$. This leads us to the renormalized moments

$$\begin{aligned}
&p\left(N, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) \\
&\quad= 1 - \frac{\alpha_s}{2\pi}\Bigg[n_f T_F\left(\frac{2}{3}\log\frac{\mu^2}{k_\perp^2} - \frac{1}{3}\right) - 4C_A\sum_{J=1}^N\binom{N}{J}(-1)^J\Bigg(\mathrm{B}_{1/2}(J,0) - 2\,\mathrm{B}_{1/2}(J+1,0) \\
&\qquad\qquad + 3\,\mathrm{B}_{1/2}(J+2,0) - 2\,\mathrm{B}_{1/2}(J+3,0) + \mathrm{B}_{1/2}(J+4,0)\Bigg)\log\frac{\mu^2}{k_\perp^2}\Bigg]\,,
\end{aligned} \tag{5.35}$$

where $\mathrm{B}_x(a,b)$ is the incomplete Euler beta function.[2] This has anomalous dimension

$$\begin{aligned}
&\gamma\left(N, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\\
&\quad= \gamma\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) + 4\frac{\alpha_s C_A}{\pi}\sum_{J=1}^{N}\binom{N}{J}(-1)^J\Big(\mathrm{B}_{1/2}(J,0) - 2\,\mathrm{B}_{1/2}(J+1,0)\\
&\qquad\qquad + 3\,\mathrm{B}_{1/2}(J+2,0) - 2\,\mathrm{B}_{1/2}(J+3,0) + \mathrm{B}_{1/2}(J+4,0)\Big)
\end{aligned} \tag{5.37}$$

Finally, let us consider the conditional probability for a gluon in the UV to split into a quark in the IR. This requires a resolved emission

$$\begin{aligned}
&p_\epsilon^{(\rm R)}\left(z_{\rm w}, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{q\bar{q}\leftarrow g}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\Theta(z>1/2)\,\delta(z-z_{\rm w})\,,\\
&\quad= \frac{\alpha_s}{2\pi}\frac{\left(\mu^2 e^{\gamma_E}\right)^\epsilon}{\Gamma(1-\epsilon)}\int_{k_\perp^2}^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{q\bar{q}\leftarrow g}\left(z_{\rm w},\ell_\perp\right)\Theta(z_{\rm w}>1/2)\,,
\end{aligned} \tag{5.38}$$

where, this being the sole contribution at this order, we have

$$p_\epsilon\left(z_{\rm w}, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) = p_\epsilon^{(\rm R)}\left(z_{\rm w}, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,. \tag{5.39}$$

The present case being a flavor-changing channel, no problems are encountered in the evaluation of moments

$$p_\epsilon\left(N, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) = \int_0^1 dz_{\rm w}\, z_{\rm w}^N p_\epsilon\left(z_{\rm w}, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,. \tag{5.40}$$

The renormalized zeroth moment is then

$$p\left(0, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) = \frac{\alpha_s T_F}{2\pi}\left(\frac{1}{3}\log\frac{\mu^2}{k_\perp^2} - \frac{1}{6}\right)\,, \tag{5.41}$$

where the anomalous dimension is given by

$$\gamma\left(0, q_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) = \frac{1}{3}\frac{\alpha_s T_F}{\pi}\,. \tag{5.42}$$

---

[2]The incomplete Euler Beta function is defined by

$$\mathrm{B}_x(a,b) \equiv \int_0^x dt\, t^{a-1}(1-t)^{b-1}\,. \tag{5.36}$$

Note that we obtain identical expressions for the gluon splitting into an anti-quark

$$p\left(0, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) = p\left(0, \bar{q}_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) \,,$$
$$\gamma\left(0, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) = \gamma\left(0, \bar{q}_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) \,. \tag{5.43}$$

At this stage, we can confirm the sum rule that results from a gluon in the UV

$$1 = p\left(0, g_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) + 2n_f\, p\left(0, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) \,,$$
$$0 = \gamma\left(0, g_{\text{ir}}(k_\perp) \,|\, g_{\text{uv}}(p_{\perp J})\right) + 2n_f\, \gamma\left(0, q_{\text{ir}}(k_\perp) \,|\, g_{\text{uv}}(p_{\perp J})\right) \,, \tag{5.44}$$

where we have made use of Eq. (5.43) to shorten the resultant sum over IR flavors. Eq. (5.44) is consistent with the general unitarity constraint of Eq. (5.23).

The renormalized expression for the higher moments is then

$$p\left(N, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right)$$
$$= \frac{\alpha_s T_F}{2\pi}\left(\frac{2^{-N-2}\left(2^{N+2}N(N+3) + 2^{N+4} - N(N+5) - 8\right)}{(N+1)(N+2)(N+3)} \log\frac{\mu^2}{k_\perp^2} + \frac{2^{-N-2}\left(N - 2^{N+3} + 4\right)}{(N+2)(N+3)}\right), \tag{5.45}$$

with anomalous dimension

$$\gamma\left(N, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) = \frac{\alpha_s T_F}{\pi}\left(\frac{2^{-N-2}\left(2^{N+2}N(N+3) + 2^{N+4} - N(N+5) - 8\right)}{(N+1)(N+2)(N+3)}\right) . \tag{5.46}$$

In this section, we have considered the joint distributions in the energy fraction of the WTA axis and IR flavor conditional on the UV flavor of the parton initiating a given jet. In evaluating the moments in z_{w}, we have obtained anomalous dimensions which are Mellin moments of Altarelli-Parisi splitting functions of Eq. (5.1), subject to the $\mathcal{O}(\alpha_s)$ WTA constraint limiting eneergy fraction to take on values $z \in [1/2, 1]$.

5.2.2 Calculation for processes involving heavy quark flavors

This section repeats much of the same analysis of Sec. 5.2.1, but with the replacement of all light quark flavors by their heavy counterparts. As mentioned previously, doing so induces non-trivial m_Q-dependence in the expressions for z_{w}-moments, but leaves all anomalous dimensions

unchanged. All instances of m_Q originate from the splitting functions involving heavy quarks:

$$\begin{aligned}
\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right) &= \frac{C_F}{\ell_\perp^2+(1-z)^2m_Q^2}\left(\frac{1+z^2}{1-z}-\epsilon\,(1-z)-\frac{2m_Q^2\,z(1-z)}{\ell_\perp^2+(1-z)^2m_Q^2}\right)\,,\\
\hat{P}_{gQ\leftarrow Q}\left(z,\ell_\perp\right) &= \frac{C_F}{\ell_\perp^2+z^2m_Q^2}\left(\frac{1+(1-z)^2}{z}-\epsilon\,z-\frac{2m_Q^2\,z(1-z)}{\ell_\perp^2+z^2m_Q^2}\right)\,,\\
\hat{P}_{Q\bar{Q}\leftarrow g}\left(z,\ell_\perp\right) &= \frac{T_F}{\ell_\perp^2+m_Q^2}\left(z^2+(1-z)^2-2\epsilon\,\frac{z(1-z)\ell_\perp^2}{\ell_\perp^2+m_Q^2}\right)\,.
\end{aligned} \tag{5.47}$$

In the evaluation of contributions to conditional probabilities originating from resolved and unresolved splittings, all massive cases can be obtained through the appropriate substitution of Eq. (5.47) in the corresponding massless cases. However, a distinctive feature pertaining to flavor-channels involving heavy quarks is that the presence of the scale m_Q leads to non-vanishing virtual contributions. We will label these with a superscript "(V)" to distinguish them from those from unresolved splittings.

Let us start with a heavy quark in the UV splitting to a heavy quark in the IR. The virtual contribution to this process is

$$\begin{aligned}
&p_\epsilon^{(\mathrm{V})}\left(z_{\mathrm{w}},Q_{\mathrm{ir}}(k_\perp)\,\middle|\,Q_{\mathrm{uv}}(p_{\perp J})\right)\\
&\quad=\delta(1-z_{\mathrm{w}})\frac{\alpha_s C_F}{2\pi}\left(\frac{1}{\epsilon^2}+\frac{1}{2\epsilon}+\frac{1}{\epsilon}\log\frac{\mu^2}{m_Q^2}+\frac{1}{2}\log^2\frac{\mu^2}{m_Q^2}+\frac{1}{2}\log\frac{\mu^2}{m_Q^2}+2+\frac{\pi^2}{12}\right)\,,
\end{aligned} \tag{5.48}$$

where the support structure defined by $\delta(1-z_{\mathrm{w}})$ reflects the fact that the emission and re-absorption of the virtual gluon leaves the heavy quark trajectory unaffected, thereby persisting to the IR to define the WTA axis. The resolved and unresolved contributions are obtained in the aforementioned fashion. For resolved emissions, we have

$$\begin{aligned}
&p_\epsilon^{(\mathrm{R})}\left(z_{\mathrm{w}},Q_{\mathrm{ir}}(k_\perp)\,\middle|\,Q_{\mathrm{uv}}(p_{\perp J})\right)\\
&\quad=\frac{\alpha_s}{2\pi}\frac{(\mu^2e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\int_0^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp>k_\perp\right)\Theta(z>1/2)\delta(z-z_{\mathrm{w}})\,,\\
&\quad=\frac{\alpha_s}{2\pi}\frac{(\mu^2e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^\infty\frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{Qg\leftarrow Q}\left(z_{\mathrm{w}},\ell_\perp\right)\Theta\left(\ell_\perp>k_\perp\right)\Theta(z_{\mathrm{w}}>1/2)\,,
\end{aligned} \tag{5.49}$$

while unresolved emissions are evaluated according to

$$p_\epsilon^{(\mathrm{U})}\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) = \delta(1-z_{\mathrm{w}})\,\frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz\int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{Qg\leftarrow Q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp < k_\perp\right)\,. \tag{5.50}$$

The full conditional probability is then obtained through the sum

$$\begin{aligned} p_\epsilon\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) &= \delta(1-z_{\mathrm{w}}) + p_\epsilon^{(\mathrm{V})}\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) + p_\epsilon^{(\mathrm{R})}\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) \\ &\quad + p_\epsilon^{(\mathrm{U})}\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right)\,. \end{aligned} \tag{5.51}$$

Interestingly, the distribution for this flavor channel, as well as all others involving non-zero quark masses, can be expressed as a sum of the corresponding massless case with an additional contribution which vanishes in the $m_Q \to 0$ limit:

$$p_\epsilon\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) = p_\epsilon\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp)\,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right) + \mathcal{I}\left(z_{\mathrm{w}}, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right)\,, \tag{5.52}$$

where we collect the expressions for the massive integrals $\mathcal{I}\left(z_{\mathrm{w}}, f_{\mathrm{ir}}(k_\perp)\,\middle|\, f_{\mathrm{uv}}(p_{\perp J})\right)$ in the appendix. We note that all ϵ-dependence is contained in $p_\epsilon\left(z_{\mathrm{w}}, q_{\mathrm{ir}}(k_\perp)\,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right)$, which is to say that the heavy-flavor moments share an identical RG structure with their light-flavored counterparts.

In the evaluation of moments, we carry over the obvious notation for the massive integrals, that is

$$\mathcal{I}\left(N, f_{\mathrm{ir}}(k_\perp)\,\middle|\, f_{\mathrm{uv}}(p_{\perp J})\right) \equiv \int_0^1 dz_{\mathrm{w}}\, z_{\mathrm{w}}^N\, \mathcal{I}\left(z_{\mathrm{w}}, f_{\mathrm{ir}}(k_\perp)\,\middle|\, f_{\mathrm{uv}}(p_{\perp J})\right)\,. \tag{5.53}$$

This tells us that the zeroth moment is

$$p\left(0, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right) = p\left(0, q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) + \mathcal{I}\left(0, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right)\,, \tag{5.54}$$

which accordingly inherits the anomalous dimension

$$\gamma\left(0, Q_{\mathrm{ir}}(k_\perp)\,\middle|\, Q_{\mathrm{uv}}(p_{\perp J})\right) = \gamma\left(0, q_{\mathrm{ir}}(k_\perp)\,\middle|\, q_{\mathrm{uv}}(p_{\perp J})\right)\,. \tag{5.55}$$

The computation of higher moments follows the same steps as in the light-flavor case, involving the use of what we refer to as the "trick" of Eq. (5.11). The final result is simply a generalization of the structure observed above, namely

$$p_\epsilon\left(N, Q_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = p_\epsilon\left(N, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(N, Q_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right)\,, \quad (5.56)$$

so that the renormalized distribution is

$$p\left(N, Q_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = p\left(N, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(N, Q_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right)\,, \quad (5.57)$$

with anomalous dimension

$$\gamma\left(N, Q_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = \gamma\left(N, q_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\,. \quad (5.58)$$

Next, let us consider the conditional probability for a heavy quark in the UV to split to a gluon in the IR. Like all other flavor-changing channels, this requires a resolved splitting

$$\begin{aligned} &p_\epsilon^{(\rm R)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) \\ &\quad= \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gQ\leftarrow Q}\left(z,\ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\Theta(z>1/2)\,\delta(z-z_{\rm w})\,, \\ &\quad= \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_{k_\perp^2}^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{gQ\leftarrow Q}\left(z_{\rm w},\ell_\perp\right)\Theta(z_{\rm w}>1/2)\,. \end{aligned} \quad (5.59)$$

Thus, we have

$$p_\epsilon^{(\rm R)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = p_\epsilon^{(\rm R)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right)\,, \quad (5.60)$$

so that

$$p_\epsilon\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = p_\epsilon\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right)\,. \quad (5.61)$$

The renormalized zeroth moment is then

$$p\left(0, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = p\left(0, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(0, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right)\,, \quad (5.62)$$

and has anomalous dimension

$$\gamma\left(0, g_{\rm ir}(k_\perp)\,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = \gamma\left(0, g_{\rm ir}(k_\perp)\,\middle|\, q_{\rm uv}(p_{\perp J})\right)\,, \quad (5.63)$$

and thus RG-consistency is automatically satisfied upon the inclusion of heavy flavors in the IR. We further confirm that the probabilistic sum rule is maintained under the addition of heavy flavors. This latter point is equivalent to the following sum rule for the mass-dependent integrals

$$0 = \mathcal{I}\left(0, Q_{\rm ir}(k_\perp) \,\middle|\, Q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(0, g_{\rm ir}(k_\perp) \,\middle|\, Q_{\rm uv}(p_{\perp J})\right) \,. \tag{5.64}$$

For general moments, the result follows the expected pattern:

$$p\left(N, g_{\rm ir}(k_\perp) \,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = p\left(N, g_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(N, g_{\rm ir}(k_\perp) \,\middle|\, Q_{\rm uv}(p_{\perp J})\right) \,, \tag{5.65}$$

where the anomalous dimensions are given by

$$\gamma\left(N, g_{\rm ir}(k_\perp) \,\middle|\, Q_{\rm uv}(p_{\perp J})\right) = \gamma\left(N, g_{\rm ir}(k_\perp) \,\middle|\, q_{\rm uv}(p_{\perp J})\right) \,. \tag{5.66}$$

Next, we analyze the cases originating from a gluon in the UV. We begin with the non-flavor-changing channel. As in the heavy-quark-to-heavy-quark transition, that of gluon-to-gluon receives a non-zero virtual correction from the self-energy diagram involving a heavy-quark loop. This virtual contribution is given by

$$p_\epsilon^{(\rm V)}\left(z_{\rm w}, g_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) = -\frac{2}{3}\frac{\alpha_s n_Q T_F}{2\pi}\left(\frac{1}{\epsilon} + \ln\frac{\mu^2}{m_Q^2}\right) \,, \tag{5.67}$$

for n_Q heavy-quark flavors.

The resolved contribution is the same as in Eq. (5.27), the unresolved component of Eq. (5.28) requires modification—that is to include the unresolved splitting of heavy quarks in addition to light quarks

$$\begin{aligned}
&p_\epsilon^{(\rm U)}\left(z_{\rm w}, g_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) \\
&\quad = \delta(1-z_{\rm w})\,\frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\Big[\hat{P}_{gg\leftarrow g}(z,\ell_\perp) + n_f \hat{P}_{q\bar{q}\leftarrow g}(z,\ell_\perp) \\
&\qquad\qquad + n_Q \hat{P}_{Q\bar{Q}\leftarrow g}(z,\ell_\perp)\Big]\Theta(\ell_\perp < k_\perp) \,, \\
&\quad = \delta(1-z_{\rm w})\,\frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^{k_\perp^2} \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\Big[\hat{P}_{gg\leftarrow g}(z,\ell_\perp) + n_f \hat{P}_{q\bar{q}\leftarrow g}(z,\ell_\perp) \\
&\qquad\qquad + n_Q \hat{P}_{Q\bar{Q}\leftarrow g}(z,\ell_\perp)\Big] \,.
\end{aligned} \tag{5.68}$$

The total conditional probability for a gluon in the UV to retain its identity in the IR is then

$$\begin{aligned}
&p_\epsilon\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\\
&\quad= \delta(1-z_{\rm w}) + p_\epsilon^{(\rm V)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) + p_\epsilon^{(\rm R)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\\
&\quad+ p_\epsilon^{(\rm U)}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,,
\end{aligned} \tag{5.69}$$

so that we have

$$\begin{aligned}
&p_\epsilon\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\Big|_{\text{with heavy flavors}}\\
&\quad= p_\epsilon\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\Big|_{\text{without heavy flavors}} + \mathcal{I}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,,
\end{aligned} \tag{5.70}$$

where we have forfeited succinct notation in favor of disambiguation. It is worth noting that $\mathcal{I}\left(z_{\rm w}, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)$ originates purely from the unresolved splitting, and thus proportional to $\delta(1-z_{\rm w})$.

The inclusion of heavy quark effects in the gluon-to-gluon channel are thus as one might reasonably expect. That is, the addition of massive integral and the replacement of instances of n_f with $n_f + n_Q$, as can bee seen in the renormalized zeroth moment

$$p\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right) = 1 - \frac{\alpha_s(n_f+n_Q)T_F}{2\pi}\left(\frac{2}{3}\log\frac{\mu^2}{k_\perp^2} - \frac{1}{3}\right) + \mathcal{I}\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,, \tag{5.71}$$

whose anomalous dimension is then modified to

$$\gamma\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\Big|_{\text{with heavy flavors}} = -\frac{2}{3}\frac{\alpha_s(n_f+n_Q)T_F}{\pi}\,, \tag{5.72}$$

which is, of course, identical to Eq. (5.32), only requiring accounting for the increased number active quark flavors.

This leads us to the higher moments, whose renormalized expressions are then

$$\begin{aligned}
&p\left(N, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\\
&\quad= 1 - \frac{\alpha_s}{2\pi}\Bigg[(n_f+n_Q)T_F\left(\frac{2}{3}\log\frac{\mu^2}{k_\perp^2} - \frac{1}{3}\right) + \mathcal{I}\left(0, g_{\rm ir}(k_\perp)\,\middle|\, g_{\rm uv}(p_{\perp J})\right)\\
&\quad- 4C_A\sum_{J=1}^{N}\binom{N}{J}(-1)^J\Big(\mathrm{B}_{1/2}(J,0) - 2\,\mathrm{B}_{1/2}(J+1,0) + 3\,\mathrm{B}_{1/2}(J+2,0)\\
&\quad- 2\,\mathrm{B}_{1/2}(J+3,0) + \mathrm{B}_{1/2}(J+4,0)\Big)\log\frac{\mu^2}{k_\perp^2}\Bigg]\,,
\end{aligned} \tag{5.73}$$

where we note that $\mathcal{I}\left(0, g_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)$ shows up due to the presence of unresolved-heavy-quark splittings, which in turn emerges from the calculational trick of Eq. (5.11). These moments evolve according to the anomalous dimension

$$\begin{aligned}
&\gamma\left(N, g_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) \\
&\quad = \gamma\left(0, g_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)\Big|_{\text{with heavy flavors}} + 4\frac{\alpha_s C_A}{\pi}\sum_{J=1}^{N}\binom{N}{J}(-1)^J\Big(\mathrm{B}_{1/2}(J,0) \\
&\qquad\qquad - 2\,\mathrm{B}_{1/2}(J+1,0) + 3\,\mathrm{B}_{1/2}(J+2,0) - 2\,\mathrm{B}_{1/2}(J+3,0) + \mathrm{B}_{1/2}(J+4,0)\Big)\,,
\end{aligned} \tag{5.74}$$

which is identical to Eq. (5.37) upon taking $n_f \to n_f + n_Q$ in $\gamma\left(N, g_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)$.

Lastly, we consider the conditional probability for a gluon in the UV to split to a heavy quark in the IR. The requisite resolved splitting is

$$\begin{aligned}
&p_\epsilon^{(\rm R)}\left(z_{\rm w}, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) \\
&\quad = \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_0^1 dz \int_0^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{Q\bar{Q}\leftarrow g}\left(z, \ell_\perp\right)\Theta\left(\ell_\perp > k_\perp\right)\Theta(z > 1/2)\,\delta(z - z_{\rm w})\,, \\
&\quad = \frac{\alpha_s}{2\pi}\frac{(\mu^2 e^{\gamma_E})^\epsilon}{\Gamma(1-\epsilon)}\int_{k_\perp^2}^\infty \frac{d\ell_\perp^2}{\ell_\perp^{2\epsilon}}\,\hat{P}_{Q\bar{Q}\leftarrow g}\left(z_{\rm w}, \ell_\perp\right)\Theta(z_{\rm w} > 1/2)\,,
\end{aligned} \tag{5.75}$$

which naturally leads to

$$p_\epsilon\left(z_{\rm w}, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) = p_\epsilon^{(\rm R)}\left(z_{\rm w}, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,, \tag{5.76}$$

which can, at this point rather unsurprisingly, be expressed as

$$p_\epsilon\left(z_{\rm w}, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) = p_\epsilon\left(z_{\rm w}, q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(z_{\rm w}, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,. \tag{5.77}$$

Its renormalized zeroth moment is

$$p\left(0, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) = p\left(0, q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) + \mathcal{I}\left(0, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,, \tag{5.78}$$

which has anomalous dimension

$$\gamma\left(0, Q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right) = \gamma\left(0, q_{\rm ir}(k_\perp) \,\middle|\, g_{\rm uv}(p_{\perp J})\right)\,, \tag{5.79}$$

so that the previous sum rules are modified to the following

$$1 = p\left(0, g_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) + 2(n_f + n_Q)\, p\left(0, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) ,$$
$$0 = \gamma\left(0, g_{\text{ir}}(k_\perp) \,|\, g_{\text{uv}}(p_{\perp J})\right) + 2(n_f + n_Q)\, \gamma\left(0, q_{\text{ir}}(k_\perp) \,|\, g_{\text{uv}}(p_{\perp J})\right) , \tag{5.80}$$

which are of course unmodified if we just count the heavy flavors in the total number of active flavors, i.e. by absorbing n_Q into n_f through $n_f + n_Q \to n_f$.

Finally, the higher moments take the anticipated form

$$p\left(N, Q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) = p\left(N, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) + \mathcal{I}\left(N, Q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) , \tag{5.81}$$

with anomalous dimensions

$$\gamma\left(N, Q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) = \gamma\left(N, q_{\text{ir}}(k_\perp) \,\middle|\, g_{\text{uv}}(p_{\perp J})\right) . \tag{5.82}$$

This concludes the calculation of moments and QCD evolution factors for the flavor-binned z_{w}-distributions with the inclusion of heavy-flavored quarks. At this point it is perhaps beneficial to take stock of what we have learned in our straightforward but lengthy computations. The two key takeaways can be formulated as follows. First, the heavy quark mass m_Q constitutes an IR scale, and thus should leave all renormalization properties of our moments unchanged relative to the massless-flavor cases. We find this to be the case through the identification of all anomalous dimensions affiliated heavy flavors—in both the UV and IR— as identical to those with the heavy flavors substituted for light ones. Second, and perhaps more noteworthy, we have further found that all mass effects are relegated "correction" terms $\mathcal{I}$ which depend purely on the IR scales of the problem, that is, m_Q and $k_\perp$, and who vanish in the massless limit. The insensitivity of these correction terms to any UV scale will prove important when it comes to the beahvior of the observables considered in Sec. 5.4.2.

5.3 RG evolution for marginal distributions and Bayesian inference

5.3.1 RG evolution for marginal distributions

In this section, we present the renormalization group equations that govern the evolution of the marginal distributions—that is the flavor distributions which result from marginalizing over z_{w}.

The solutions to such equations are non-trivial as they are constrained to obey the flavor sum rule given by Eq. (5.23), which we restate here in more succinct notation:

$$1 = \sum_{f_{\text{ir}}} p\left(f_{\text{ir}} \,\middle|\, f_{\text{uv}}\right) \,, \tag{5.83}$$
$$0 = \sum_{f_{\text{ir}}} \gamma\left(f_{\text{ir}} \,\middle|\, f_{\text{uv}}\right) \,.$$

In what follows, we assume the scaling of the heavy quark mass m_Q and IR cutoff $k_\perp$ with the UV scale $p_{\perp J}$ to be of the form $m_Q \sim k_\perp \sim \sqrt{k_\perp^2 + m_Q^2} << p_{\perp J}$. The massless limit accounts for the case in which $m_Q << k_\perp$. Furthermore, we assume that there are n_f flavors of massless quarks. For this configuration, the RG equations for the marginal distributions are

$$\mu \frac{\partial}{\partial \mu} p\left(f_{\text{ir}} \,\middle|\, f_{\text{uv}}\right) = \sum_{f'} p\left(f_{\text{ir}} \,\middle|\, f'\right) \, \gamma\left(f' \,\middle|\, f_{\text{uv}}\right) \,. \tag{5.84}$$

We are primarily interested in the case in which a WTA IR flavor of the jet is identified to be that of a heavy quark Q which thus restricts Eq. (5.83) to take the following matrix form

$$\begin{aligned} &\mu \frac{\partial}{\partial \mu} \begin{pmatrix} p\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu\right) \\ p\left(Q_{\text{ir}} \,\middle|\, q_{\text{uv}}; \mu\right) \\ p\left(Q_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu\right) \end{pmatrix} \\ &= \begin{pmatrix} \gamma\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right) & \gamma\left(q_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right) & \gamma\left(g_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right) \\ \gamma\left(Q_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) & \gamma\left(q_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) & \gamma\left(g_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) \\ \gamma\left(Q_{\text{ir}} \,\middle|\, g_{\text{uv}}\right) & (2n_f - 1) \times \gamma\left(q_{\text{ir}} \,\middle|\, g_{\text{uv}}\right) & \gamma\left(g_{\text{ir}} \,\middle|\, g_{\text{uv}}\right) \end{pmatrix} \begin{pmatrix} p\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu\right) \\ p\left(Q_{\text{ir}} \,\middle|\, q_{\text{uv}}; \mu\right) \\ p\left(Q_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu\right) \end{pmatrix} \,, \end{aligned} \tag{5.85}$$

where we of course have that $\gamma\left(q_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right), \gamma\left(Q_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) = 0$ at $\mathcal{O}(\alpha_s)$. As the anomalous dimensions are universal to both massless and massive cases, we may solve Eq. (5.85) through the same methods as laid out in [LN23].

5.3.2 Bayesian inference

Again, as our interests lie purely in the case of a jet identified as heavy-flavored in the IR, the relevant solutions correspond to the the $g_{\text{uv}} \to Q_{\text{ir}}$ and $Q_{\text{uv}} \to Q_{\text{ir}}$ channels. The former is given

by

$$p\left(Q_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu\right) = \frac{1}{2n_f}\Bigg[1 - p\left(g_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu_0\right)\left(\frac{\alpha_s(\mu)}{\alpha_s(\mu_0)}\right)^{\frac{2}{\beta_0}\left(C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F\right)} - \frac{p\left(g_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu_0\right)\frac{2}{3}n_f T_F + p\left(g_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu_0\right) C_F\left(2\log 2 - \frac{5}{8}\right)}{C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F} \times \left(1 - \left(\frac{\alpha_s(\mu)}{\alpha_s(\mu_0)}\right)^{\frac{2}{\beta_0}\left(C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F\right)}\right)\Bigg], \tag{5.86}$$

while the latter is

$$\begin{aligned} p\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu\right) = \frac{1}{2n_f}\Bigg[&\frac{C_F\left(2\log 2 - \frac{5}{8}\right)}{C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F}\left(1 - \left(\frac{\alpha_s(\mu)}{\alpha_s(\mu_0)}\right)^{\frac{2}{\beta_0}\left(C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F\right)}\right) \\ &+ \frac{C_F\left(2\log 2 - \frac{5}{8}\right)}{C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F}\left(p\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu_0\right) + p\left(g_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu_0\right)\right) \\ &\times \left(\frac{\alpha_s(\mu)}{\alpha_s(\mu_0)}\right)^{\frac{2}{\beta_0}\left(C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F\right)} + (2n_f - 1)p\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu_0\right)\left(\frac{\alpha_s(\mu)}{\alpha_s(\mu_0)}\right)^{\frac{2}{\beta_0}\left(C_F\left(2\log 2 - \frac{5}{8}\right)\right)} \\ &+ \frac{p\left(Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}; \mu_0\right)\frac{2}{3}n_f T_F - p\left(g_{\text{ir}} \,\middle|\, g_{\text{uv}}; \mu_0\right) C_F\left(2\log 2 - \frac{5}{8}\right)}{C_F\left(2\log 2 - \frac{5}{8}\right) + \frac{2}{3}n_f T_F}\Bigg]. \end{aligned} \tag{5.87}$$

Our evolution proceeds from the UV to the IR, and we thus take $\mu_0 = p_{\perp J}$ and $\mu = \sqrt{k_\perp^2 + m_Q^2}$.

Now, note that $p\left(f_{\text{ir}} \,\middle|\, f_{\text{uv}}\right)$ is a distribution in a random categorical variable, f_{ir}, conditioned on another, f_{uv}. This being the case, we can apply Bayes' theorem to invert this conditional relation and obtain the distribution in UV flavor conditioned on the IR. Doing so allows us to perform the inference problem that is most compelling from an experimental standpoint. In experiment, we measure the flavor of a jet based on the WTA definition, and, based on this information captured in the IR, can then infer the flavor of the individual UV parton from whence the multi-particle jet originated. This arises through use of the chain rule for probabilities

$$p\left(f_{\text{uv}} \mid f_{\text{ir}}\right) p\left(f_{\text{ir}}\right) = p\left(f_{\text{uv}}, f_{\text{ir}}\right) = p\left(f_{\text{ir}} \mid f_{\text{uv}}\right) p\left(f_{\text{uv}}\right) , \tag{5.88}$$

from which we obtain

$$p\left(f_{\text{uv}} \mid f_{\text{ir}}\right) = \frac{p\left(f_{\text{ir}} \mid f_{\text{uv}}\right) p\left(f_{\text{uv}}\right)}{p\left(f_{\text{ir}}\right)} , \tag{5.89}$$

where the denominator can be understood to arise through marginalizing over f_{uv}:

$$p\left(f_{\text{ir}}\right) = \sum_{f_{\text{uv}}} p\left(f_{\text{ir}}, f_{\text{uv}}\right) = \sum_{f_{\text{uv}}} p\left(f_{\text{ir}} \mid f_{\text{uv}}\right) p\left(f_{\text{uv}}\right) . \tag{5.90}$$

Thus, to reiterate, at this stage we are in a position to evaluate the distribution $p\left(f_{\rm uv} \,|\, f_{\rm ir}\right)$, which is to say that given the measurement of a jet at the IR scale $k_\perp$ of flavor $f_{\rm ir}$, what is the probability that this jet was initiated by a parton of flavor $f_{\rm uv}$ emerging out of the hard scattering event of the collision. The conditional probabilities, $p\left(f_{\rm ir} \,\middle|\, f_{\rm uv}\right)$, appearing in Eqs. (5.89) and (5.90) are computed according to Sec. 5.2, while the prior probabilities, $p\left(f_{\rm uv}\right)$, may be obtained though Madgraph [Alw08, AFF14] combined with PYTHIA [Bie22], simply as the differential cross sections to produce a hard parton of flavor $f_{\rm uv}$ with transverse momentum $p_{\perp J}$, and over some experimentally-relevant rapidity range, that is $d\sigma_{f_{\rm uv}}/d\eta\, dp_{\perp J}$.

Once this is performed, the posterior distribution of Eq. (5.89) takes the following form:

$$\begin{aligned} p\left(Q_{\rm uv} \,|\, Q_{\rm ir}\right) &= \frac{p\left(Q_{\rm ir} \,|\, Q_{\rm uv}\right) p\left(Q_{\rm uv}\right)}{p\left(Q_{\rm ir} \,|\, Q_{\rm uv}\right) p\left(Q_{\rm uv}\right) + p\left(Q_{\rm ir} \,|\, g_{\rm uv}\right) p\left(g_{\rm uv}\right)}\,, \\ p\left(g_{\rm uv} \,|\, Q_{\rm ir}\right) &= \frac{p\left(Q_{\rm ir} \,|\, g_{\rm uv}\right) p\left(g_{\rm uv}\right)}{p\left(Q_{\rm ir} \,|\, Q_{\rm uv}\right) p\left(Q_{\rm uv}\right) + p\left(Q_{\rm ir} \,|\, g_{\rm uv}\right) p\left(g_{\rm uv}\right)}\,, \end{aligned} \tag{5.91}$$

where the first equation is the probability for the Q-flavored jet measured in the IR to have its progenitor as the same heavy quark Q in the UV, while the second is the probability that it proceeded from a hard gluon. The resulting posterior probabilities for bottom and charm jets for both sPHENIX and LHC kinematics are displayed in Fig. 5.1. If we hold the kinematics fixed and compare the bottom case to that of charm, we see the behavior of the posteriors to depend rather mildly on the mass of the quark, considering $m_b \approx 4.18$ GeV and $m_c \approx 1.27$ GeV. The only additional feature that differentiates the cases would be the relative magnitudes of the hard-scattering cross sections. The ratios for the gluonic and heavy quark cross sections at sPEHNIX kinematics are

$$\left.\frac{p(g_{\rm uv})}{p(b_{\rm uv})}\right|_{sPHENIX} \approx 6.8\times 10^2\,, \qquad \left.\frac{p(g_{\rm uv})}{p(c_{\rm uv})}\right|_{sPHENIX} \approx 6.2\times 10^2\,, \tag{5.92}$$

while those for LHC kinematics take on values

$$\left.\frac{p(g_{\rm uv})}{p(b_{\rm uv})}\right|_{LHC} \approx 8.1\times 10^2\,, \qquad \left.\frac{p(g_{\rm uv})}{p(c_{\rm uv})}\right|_{LHC} \approx 7.2\times 10^2\,. \tag{5.93}$$

We thus see that the numerical differences observed in going from bottom to charm are quite mild as far as the prior probabilities go. In going from sPHENIX to LHC kinematics, we see that the relative contributions of heavy quarks and gluons in the UV are comparable, thus this transition mainly captures the relative effects of RG evolution, or equivalently, the effects of the radiative

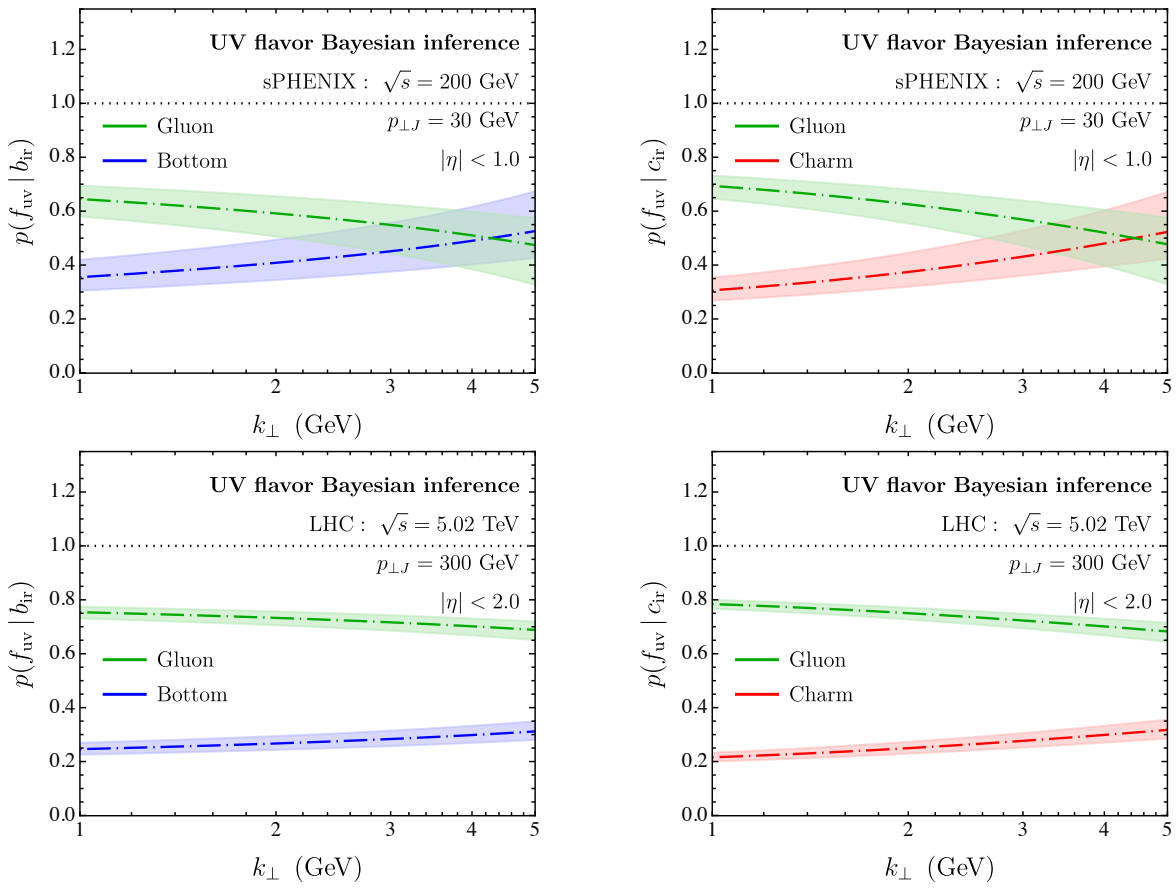

Figure 5.1: The posterior probabilities for a particular UV flavor sourcing a bottom quark in the IR. Kinematics relevant for sPHENIX (above) and the LHC (below) are displayed.

partonic cascade. This is due simply to the larger phase space available to jets measured at the LHC relative to those at sPEHNIX. We see that in both the cases of bottom and charm, by increasing the UV scale of the jet, and thus widening the regime of QCD evolution, the posterior probability for a gluon to source the Q-flavored jet is increased. This may be understood intuitively, as the more phase space available to a given jet for its subsequent radiation pattern, the higher chance there is for a gluon to split into a $Q\bar{Q}$ pair at some point in its evolution.

Thus for sPHENIX kinematics if one measures a Q-flavored jet in the IR, the chance that it originated from a hard gluon in the UV is roughly that of a coin flip across the range of $k_\perp \sim$ 1-5 GeV. This then reaches the $\sim$ 75-80% level if one measures such a jet at the LHC. Therefore, we can attribute an approximately 10-20% increase in the gluonic sourcing of heavy-quark jets

to the added complexity of parton showers probed at the LHC. Now, it is important to note that the posterior probabilities displayed in Fig. 5.1 are those that result in measuring a heavy-quark flavored jet under no special circumstances, i.e. under no particular phase-space restrictions beyond falling within basic rapidity window cited in the plots. In other words, these probabilities are those derived under the most agnostic of experimental cuts. These probabilities certainly depend on the context in which a heavy-quark jet is measured, which is to say, they have non-trivial dependence on the underlying process under experimental consideration.

With these considerations in mind, one may perform a more general analysis to estimate how a strongly a particular process which results in the measurement of heavy-flavored jets in the IR can be traced back to heavy quarks in the UV. We will refer to such a probability as the IR to UV heavy quark fidelity, which will be taken as the posterior probability that $Q_{\rm uv}$ had given rise to the measured $Q_{\rm ir}$. Let us simplify the expression for this probability as given by Eq. (5.91) by pulling out a factor of $p\left(Q_{\rm ir} \,|\, Q_{\rm uv}\right) p\left(Q_{\rm uv}\right)$. For further simplification, we define the following ratio

$$\Phi \equiv \frac{p(g_{\rm uv})}{p(Q_{\rm uv})}\,, \tag{5.94}$$

so that the IR to UV heavy quark fidelity takes the suucint form

$$p\left(Q_{\rm uv} \,|\, Q_{\rm ir}; \Phi\right) = \left(1 + \Phi\, \frac{p\left(Q_{\rm ir} \,|\, g_{\rm uv}\right)}{p\left(Q_{\rm ir} \,|\, Q_{\rm uv}\right)}\right)^{-1}\,. \tag{5.95}$$

In the above form, we are treating the heavy-quark fidelity as a one-parameter family of distributions, parameterized by Φ. This parameter carries all dependence on the kinematics of the collision system, as well as that due to the physical process under consideration. Of course, the ratio of conditional probabilities appearing in Eq. (5.95) depends on the kinematics and process under consideration to the extent that these factors lead to the hard scale $p_{\perp J}$ on which both $f_{\rm uv}$ and the QCD radiation pattern implicitly depend—but Φ is the most direct and manifest probe.

The behavior of Eq. (5.95) under a wide range of physical conditions may be gleaned through inspection of Fig. 5.2. In the upper two panels, we have the ratios of conditional probabilities appearing in Eq. (5.95), that is the ratio of the probability for the heavy quark in the IR to be sourced by a gluon in the UV to that of a heavy quark in the UV—the case of bottom flavor on the left and charm on the right. The effects of QCD evolution under a wide kinematic range are

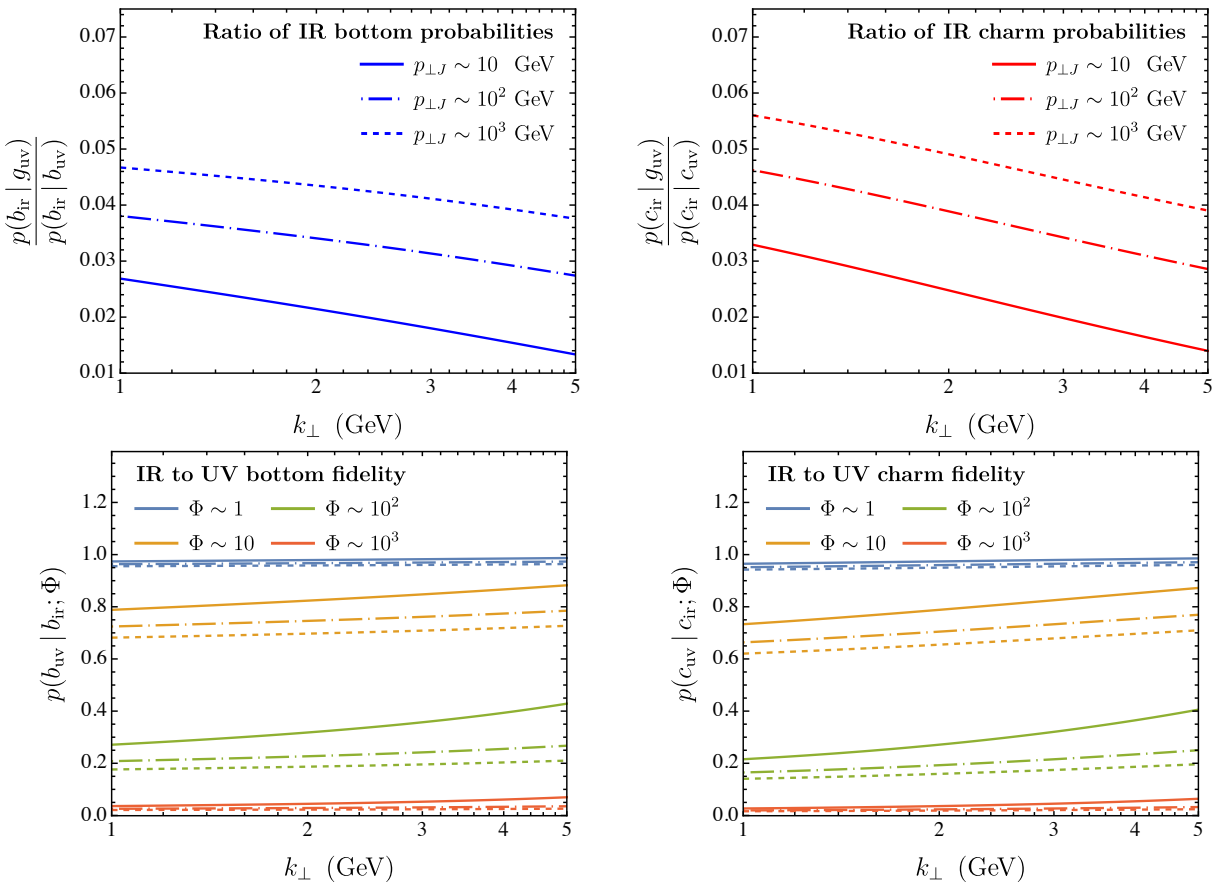

Figure 5.2: (Above) Ratio of conditional probabilities appearing in the definition of the heavy quark fidelity for various levels of RG-running. (Below) Heavy quark fidelity for various ratios Φ of the probabilities to produce a gluon vs. a heavy quark in the UV. The line dashing in the lower plots follows the same convention as signified in the upper plots.

displayed by the three curves in each plot. On this front, we get a better understanding of the pattern revealed by Fig. 5.1 in going from sPHENIX to LHC kinematics—that is the lower the UV scale $p_{\perp J}$, which determines the phase space availability for subsequent parton shower, the smaller the relative size of gluonic sourcing of heavy quarks, and vice versa. Although we observe a slightly less than doubling of the ratios through two orders of magnitude in UV-scale variation, for both bottom and charm flavors, it is important to note that these ratios remain at the level of about a few parts-per-hundred.

In the lower two panels of Fig. 5.2 we display the IR to UV heavy-quark fidelity, again with

bottom shown in the left panel and charm in the right. In these plots, we have four triplets of curves, where each triplet corresponds to a particular value of the parameter Φ, which varies based on the process under consideration. Within each triplet, the line-dashing corresponds to the values of $p_{\perp J}$ as specified in the upper two panels. The two extreme limits can be immediately intuited—for we must have $p\left(Q_{\mathrm{uv}} \,\middle|\, Q_{\mathrm{ir}}; \Phi\right) \to 1$ as $\Phi \to 1$, since the conditional probability ratio is $\mathcal{O}(10^{-2})$. Conversely, as $\Phi \to 10^3$ we expect $p\left(Q_{\mathrm{uv}} \,\middle|\, Q_{\mathrm{ir}}; \Phi\right) \to 0$ since this limit makes the product of Φ with the conditional probability ratio $\mathcal{O}(10)$. Between these two extremes lies the transition, where $\Phi \sim 10^2$ yields a low heavy quark fidelity, and $\Phi \sim 10$ results in a rather high one. Given this context, we can understand the relative placements of the heavy-flavor curves at sPHENIX and LHC kinematics in Fig. 5.1. In both of these cases, Φ lies closer to the $\mathcal{O}(10^2)$ level than $\mathcal{O}(10)$, situating the fidelity in the lower end of the spectrum, sPHENIX in the roughly 40% range and LHC more nearly 20%. The remarkable fact, though, is that for $\Phi \sim 10$ one can achieve around 70% fidelity. Thus, through careful observable design, if one can isolate physical processes where the hard gluon cross section is only a few to a several times that from heard heavy quarks, one may expect a near-90% fidelity. This is of particular interest for heavy-ion experiments in which the flavor-dependence of partonic energy loss has been a long standing question [KRV19b]. By designing observables with high IR to UV heavy-quark fidelity, one can be assured that the resulting modification pattern faithfully represents the flavor-dependence that one seeks. Furthermore, one could utilize this UV fidelity to ask questions regarding the partonic flavors involved in the initial state of the hard collision and thus infer properties regarding the initial state, such as the Sivers function [KRS21].

5.4 RG evolution for higher moments and jet shape observables

5.4.1 Evolution

In this section, we turn our attention to the evolution of the higher z_{w}-moments of our distributions. There is an important distinction to be made regarding the evolution of such moments relative to the zeroth moments/marginal distributions. The first moment is the average fractional energy carried by the WTA axis, and while at NLO it is technically constrained to lie between 1/2 and 1 (because

$z_{\rm w}$ is constrained to lie within this range), there are no such constraints involving the sum over IR flavors. For instance, with heavy quarks lying along the WTA axis, there are no limits on their average energy-fractions which relate them to those of gluons along the WTA axis, as these are fundamentally different physical situations. Such is a general feature of all the higher moments—unlike the zeroth moment, there are no sum rules like Eq. (5.83) to consider.

The first $z_{\rm w}$-moment of our conditional distribution is related to the jet shape [CLM22a], and we will carry this language over to generalize for all higher moments. Traditionally, the integrated jet shape measured about the WTA axis is defined with respect to an angular displacement measured relative to the WTA axis

$$\Psi(r) = \sum_{i\in J} z_i \,\Theta\left(r - r_{\mathrm{wta},i}\right) \,, \tag{5.96}$$

where $z_i = p_{\perp i}/p_{\perp J}$ of the $i^{\rm th}$ particle contained in the jet. This then makes the differential shape

$$\psi(r) = \frac{d}{dr}\Psi(r) = \sum_{i\in J} z_i \,\delta\left(r - r_{\mathrm{wta},i}\right) \,. \tag{5.97}$$

It is important to point out here that these definitions are made with respect to an angular scale r measured about the WTA axis. In what follows, we will compute shapes with respect to the mass-dimension-one IR resolution scale $k_\perp$ measured about the WTA axis.

Let us generalize the notion of the integrated jet shape for higher moments and with respect to $k_\perp$. This is straightforwardly done via

$$\Psi(N, k_\perp) = \sum_{i\in J} z_i^N \,\Theta\left(k_\perp - k_{\perp\mathrm{wta},i}\right) \,, \tag{5.98}$$

where again, we highlight the fact that this observable is no dependent on the resolution scale $k_\perp$. Once such observables are measured for an ensemble of jets, they are can then be related directly back to the $z_{\rm w}$-moments computed in Sec. 5.2.

To keep in line with the notation established in Sec. 5.2, these shapes are denoted as $\Psi_f(N,\mu)$, and defined as

$$\Psi_f(N,\mu) \equiv \sum_{f_{\rm ir}} p\left(N, f_{\rm ir}(k_\perp) \,\middle|\, f(p_{\perp J})\right) \,, \tag{5.99}$$

where note we are leaving the conditional flavor label general. The intuition for doing so goes as follows: we consider a fixed flavor f which gives rise to a jet whose energy-fraction moments we are ultimately measuring. As the jet evolves, energy is distributed across its various decay products, and we observe this distribution in the IR. To consider the full dynamical evolution, we thus need to consider two pieces of information—first is the fraction of jets of a particular IR flavor that stem from the fixed UV flavor. Second is the resulting moment for such an IR flavor. These two pieces of information need be combined in order to properly evolve the jet shapes of Eq. (5.99).

The RG matrix governing the evolution of these jet shapes can be interpreted as the N^{th} moment of the WTA-modified DGLAP matrix, which in our notation is given by

$$\mathbf{\Gamma}(N) = \begin{pmatrix} \gamma\left(N, Q_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right) & \gamma\left(N, Q_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) & \gamma\left(N, Q_{\text{ir}} \,\middle|\, g_{\text{uv}}\right) \\ \gamma\left(N, q_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right) & \gamma\left(N, q_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) & (2n_f - 1) \times \gamma\left(N, q_{\text{ir}} \,\middle|\, g_{\text{uv}}\right) \\ \gamma\left(N, g_{\text{ir}} \,\middle|\, Q_{\text{uv}}\right) & \gamma\left(N, g_{\text{ir}} \,\middle|\, q_{\text{uv}}\right) & \gamma\left(N, g_{\text{ir}} \,\middle|\, g_{\text{uv}}\right) \end{pmatrix}, \tag{5.100}$$

so that the RG evolution equations become

$$\mu \frac{\partial}{\partial \mu} \begin{pmatrix} \Psi_Q(N,\mu) \\ \Psi_q(N,\mu) \\ \Psi_g(N,\mu) \end{pmatrix} = \mathbf{\Gamma}(N) \begin{pmatrix} \Psi_Q(N,\mu) \\ \Psi_q(N,\mu) \\ \Psi_g(N,\mu)\,. \end{pmatrix} \tag{5.101}$$

Now, to properly account for the evolution process described in the discussion surrounding Eq. (5.99), we act evolution matrix corresponding to $\mathbf{\Gamma}(N)$ on the following initial condition

$$\begin{pmatrix} \Psi_Q(N, p_{\perp J}) \\ \Psi_q(N, p_{\perp J}) \\ \Psi_g(N, p_{\perp J}) \end{pmatrix} = \left[\sum_{i=1}^{3} \left(\frac{\alpha_s(p_{\perp J})}{\alpha_s\left(k_\perp\right)} \right)^{-\lambda_i(N)} \mathbf{P}_i(N) \right] \begin{pmatrix} \Psi_Q\left(N, k_\perp\right) p\left(Q_{\text{ir}} \,|\, f_{\text{uv}}\right) \\ \Psi_q\left(N, k_\perp\right) p\left(q_{\text{ir}} \,|\, f_{\text{uv}}\right) \\ \Psi_g\left(N, k_\perp\right) p\left(g_{\text{ir}} \,|\, f_{\text{uv}}\right) \end{pmatrix}, \tag{5.102}$$

where $\mathbf{P}_i(N)$ are the projectors onto the eigenspaces of $\mathbf{\Gamma}(N)$, namely

$$\mathbf{P}_i(N) = \frac{\left(\mathbf{\Gamma}(N) - \lambda_j \mathbf{1}\right)\left(\mathbf{\Gamma}(N) - \lambda_k \mathbf{1}\right)}{(\lambda_i - \lambda_j)(\lambda_i - \lambda_k)}\,, \tag{5.103}$$

with $\mathbf{1}$ being the 3×3 identity matrix and $i \neq j \neq k$.

Let us examine this initial condition in more detail. We fix a UV flavor f_{uv} and consider the distribution of IR-flavored jets this gives rise to. Then we take the fraction for each flavor and

use it to scale the $z_{\rm w}$-moment which results from that particular flavor in the IR. We then evolve the combined object back up to the UV scale $p_{\perp J}$. It is this requisite scaling of $\Psi_f(N, k_\perp)$ by $p\left(f \,\middle|\, f_{\rm uv}\right)$ in the initial condition vector that motivates the subscript f which adorns the jet shapes. This is a subtle point.

5.4.2 Moments/shape observables

In this section, we present phenomenological predictions for two differential observables which end up being $k_\perp$-derivatives of the first two $z_{\rm w}$-cumulants of the conditional distribution $\sum_{f_{\rm ir}} p\left(z_{\rm w}, f_{\rm ir} \mid f_{\rm uv}\right)$, by relation to Eq. (5.99). For such purposes, we will consider the restricted case where heavy quarks emerge from the hard scattering, which can be achieved through the measuring of jet shapes on back-to-back heavy-quark-tagged jet pairs. Thus, we will fix $f_{\rm uv} = Q_{\rm uv}$.

The first differential shape observable will be the $k_\perp$-derivative of the first cumulant/moment

$$\begin{aligned}\psi(k_\perp) &\equiv \frac{d}{dk_\perp}\Psi(1, k_\perp)\,, \\ &= \frac{d}{dk_\perp}\left\langle z_{\rm w}\right\rangle\,,\end{aligned} \tag{5.104}$$

which we denote by ψ due to its close structural relation to the traditional differential shape. As such, we will simply refer to this observable as the differential jet shape, keeping in mind though that its dependence on $k_\perp$ distinguishes it as novel. The second observable will be the $k_\perp$-derivative of the second cumulant, which we denote by

$$\begin{aligned}\xi(k_\perp) &\equiv \frac{d}{dk_\perp}\left(\Psi(2, k_\perp) - \Psi(1, k_\perp)^2\right)\,, \\ &= \frac{d}{dk_\perp}\left(\left\langle z_{\rm w}^2\right\rangle - \left\langle z_{\rm w}\right\rangle^2\right)\,,\end{aligned} \tag{5.105}$$

and will refer to as the differential jet dispersion.

Let us first consider the former observable, $\psi(k_\perp)$. Results for this are displayed in Fig. 5.3 where we fix $f_{\rm uv} = b$ in the left two panels and $f_{\rm uv} = c$ in the right two. The upper two panels display the resulting predictions for jets of sPHENIX kinematics while the lower two show those for the LHC. First, suppose we fix the kinematics and then examine the effects of the heavy quark mass—to do so we fix either the upper or lower panels and read from left to right. We see that

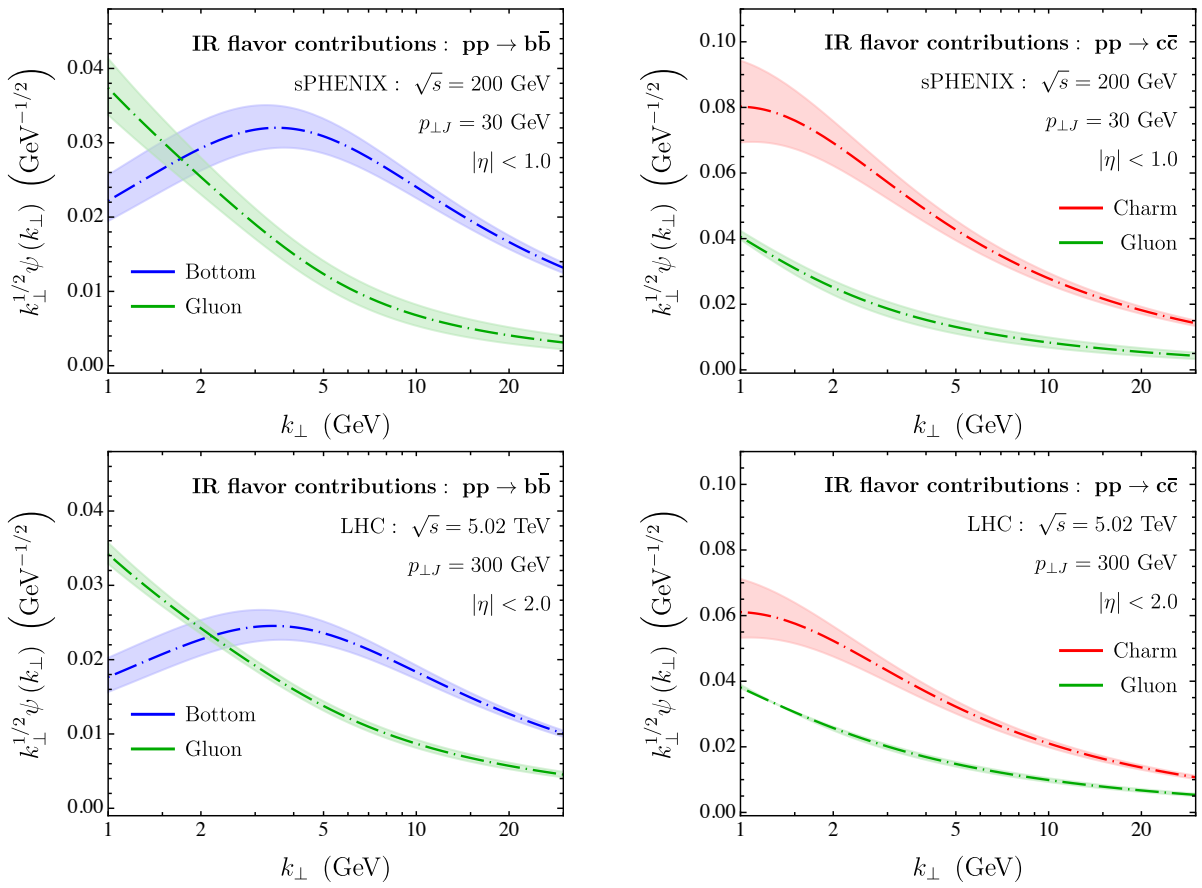

Figure 5.3: Differential jet shapes resulting from heavy quarks in the final state of the hard-scattering cross section, for both bottom (left) and charm (right) flavors. Kinematics relevant for sPHENIX (above) and the LHC (below) are displayed.

in doing so, the heavy-quark-flavor contribution to the differential first cumulant clearly attains a global maximum for $k_\perp \sim \mathcal{O}(1) \times m_Q$—thus reading from left to right, we see this peak falls near m_b for the bottom case and m_c for charm. Contrast this functional behavior with that of the gluonic contributions, whose curves exhibit an approximately power-law dependence on $k_\perp$.

Next, let us fix the flavor and examine the effects of the chosen initial UV scale. To do so, we choose either the left or right columns and read top-down, which takes us from sPHENIX to LHC kinematics. In doing so, we increase $p_{\perp J}$ by an order of magnitude, thus increasing the virtuality of the heavy quark emerging from its respective hard collision. This increase in virtuality opens up a larger volume of phase space for the ensuing parton shower to explore, thereby dilating the

energy-window over which the radiative cascade evolves into the final-state jet measured in the IR. This brings us to an important point. As can be gleaned upon inspection of Fig. 5.3, the net effect of such an increase in the UV scale is simply a decrease in the overall normalization of the curves. This is just the physically-intuitive fact that increased "duration" of the parton shower results in a wider distribution of energy-fractions amongst the increased number of final state particles forming the jet, and hence an overall lower average value taken on by that which defines the WTA axis. This then brings us to a remarkable point—which is that increasing the UV scale has no effect on the $k_\perp$-profile of the curve itself. This of course, is due to all of the IR information of each distribution being relegated to the "mass-correction" terms of the from $\mathcal{I}(z_{\rm w}, f_{\rm ir} \,\big|\, f_{\rm uv})$, whose profiles are purely functions of m_Q and $k_\perp$ only. This is to say that they depend solely on the IR scales of the problem, which are of course invariant under changes of the UV scale. Thus, these observables are able to manifest the dead-cone effect through the location of their global maxima in a way that is universal across collisional energies. This is in sharp contrast to observables which depend on an angular resolution scale, as such a scale can then only reveal the dead-cone angle $r \sim m_Q/p_{\perp J}$, and hence is directly-dependent on the UV scale of the collision system.

Let us now consider the behavior of $\xi(k_\perp)$. Results for this observable are shown in Fig. 5.3 where, again, the left two panels correspond to $f_{\rm uv} = b$ while the right two have $f_{\rm uv} = c$. Similarly, sPHENIX kinematics are used for the upper two panels while LHC kinematics are chosen for the lower two. What distinguishes the jet dispersion from the shape is that the nonlinear relationship between the dispersion and the first two moments of the $z_{\rm w}$-distribution, as given by Eq. (5.105), allows for individual flavor contributions to this observable to be negative, as can be seen in Fig. 5.3 for the heavy-flavor contributions. Just like the differential jet shape, the differential jet dispersion exhibits a universal $k_\perp$ -profile across UV scales, while its overall normalization depends on the UV scale for the same physical reasons as that of the jet shape. Coupling the normalization effects with the allowed negativity of the heavy-quark flavor contributions is what leads the crossing points of the heavy-quark curves to be non-invariant across collisional systems. However, we see that as we increase the UV scale, the small-$k_\perp$ limit of the jet dispersions also increases slightly. This important normalization effect can be understood as a dual to the decrease that the jet shape normalization experiences. As we increase the UV scale, we distribute energy-fractions across

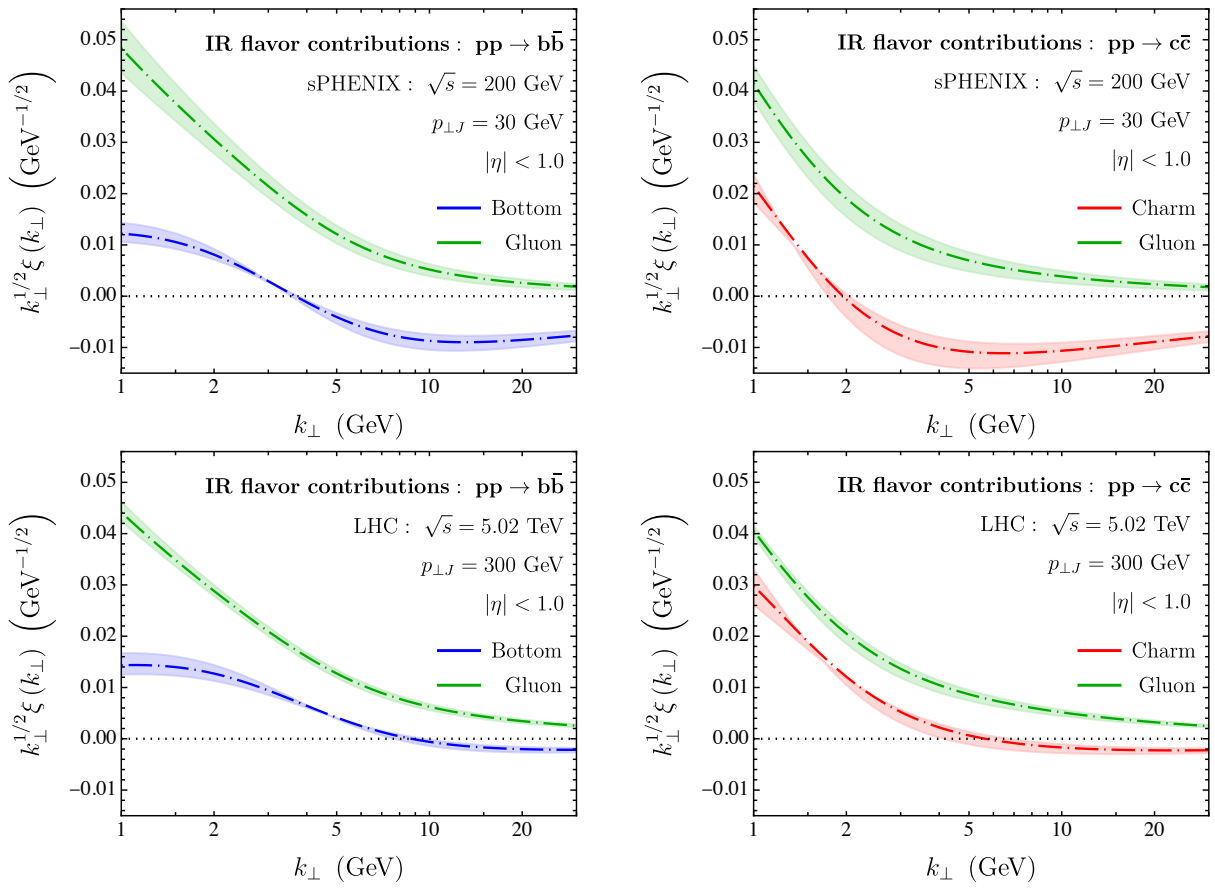

Figure 5.4: Differential jet dispersions resulting from heavy quarks in the final state of the hard-scattering cross section, for both bottom (left) and charm (right) flavors. Kinematics relevant for sPHENIX (above) and the LHC (below) are displayed.

more particles, hence lowering the average fraction carried by the WTA axis, but also increasing the variance that this value experiences. Thus, while undoubtedly simply to construct, both the differential jet shapes and dispersions are able to capture a large amount of physics.

5.5 Conclusion

In this chapter, we have extended the definition for the IR flavor for a jet proposed in [CLM22a] to encompass heavy flavors. That is, we define a jet to have a heavy flavor if that flavor constitutes the net flow along the WTA axis of the jet. With the defining characteristic of a heavy quark being

its mass m_Q, we analytically compute the m_Q-dependence that such a flavor definition induces on the distributions governing the rate of heavy flavor jet production as well as its internal substructure. We compute these distributions to NLO and find, remarkably, that at all m_Q-dependence is contained within massless-correction terms depending solely on m_Q and $k_\perp$, the resolution scale of the NLO splitting. Such dependence on purely IR scales leads to universal $k_\perp$-profiles in jet shape and dispersion observables, which are interpreted as the first two cumulants in the WTA energy fraction, $z_{\rm w}$, distributions. This leads to dead-cone effect manifesting itself in a way that is uniform across collision systems—be it collisions occurring at $\sqrt{s} = 200$ GeV at sPHENIX or $\sqrt{s} = 5.02$ TeV collisions at eh LHC. Furthermore, our construction allows for inference of the UV flavor-sourcing of a jets measured in the IR through computation of Bayesian posterior distributions. Access to such distributions empowers physicists in the design of observables that accurately probe the physics of the UV flavors that they are ultimately after. Immediate applications which come to mind include the flavor-dependence of parton energy loss in the quark-gluon plasma created in heavy-ion collisions. In order for a heavy-flavor jet to provide information regarding heavy-quark energy loss, one must be sure that the dominant contributors to those IR jets are indeed heavy quarks in the UV, created concurrently with the plasma itself. Additionally, hard heavy quarks in the UV initiating back-to-back dijet pairs stem predominantly through gluon fusion, and this fact allows for the probing of the Sivers asymmetry in deep-inelastic scattering. Thus our UV flavor inference scheme allows for the constraining of both initial and final states of the hard scattering. With such features and applications in mind, we hope that our construction leads to progress in the realm of heavy-flavor jet substructure, an exciting frontier in collider physics.

Appendix

In this appendix, we collect the explicit forms for the anomalous dimensions for the for the mass-correction terms $\mathcal{I}\left(N, f_{\rm ir} \,\middle|\, f_{\rm uv}\right)$. They are

$$\begin{aligned}\mathcal{I}\left(N, Q_{\rm ir} \,\middle|\, g_{\rm uv}\right) = T_F\Bigg(&\frac{2^{-2-N}(-4+2^{3+N}-N)}{6+5N+N^2}\frac{m_Q^2}{k_\perp^2+m_Q^2}\\ &+\frac{2^{-2-N}(8+5N+N^2)-(4+3N+N^2)}{(1+N)(2+N)(3+N)}\log\left(1+\frac{m_Q^2}{k_\perp^2}\right)\Bigg),\end{aligned} \tag{5.106}$$

$$
\begin{aligned}
\mathcal{I}\left(N, g_{\rm ir}\,\middle|\, Q_{\rm uv}\right) = C_F\Bigg(&-\frac{4}{(1+N)^2}+\frac{2}{(2+N)^2}-\frac{1}{2+N}+\frac{2^{-2-N}(32+33N+10N^2+N^3)}{(1+N)^2(2+N)^2}\\
&+\frac{k_\perp^2}{2m_Q^2}\left(-\frac{m_Q^2}{k_\perp^2}\right)^{-N/2}\left[\mathrm{Beta}\left(-\frac{m_Q^2}{k_\perp^2},2+\frac{N}{2},0\right)-\mathrm{Beta}\left(-\frac{m_Q^2}{4k_\perp^2},2+\frac{N}{2},0\right)\right]\\
&+\frac{N^2(k_\perp^2-2m_Q^2)+8m_Q^2}{k_\perp^2N(2+N)^2}{}_2F_1\left(1,1+\frac{N}{2},2+\frac{N}{2},-\frac{m_Q^2}{k_\perp^2}\right)\\
&+\frac{2^{-2-N}(-k_\perp^2N^2+2(-4+N^2)m_Q^2)}{k_\perp^2N(2+N)^2}{}_2F_1\left(1,1+\frac{N}{2},2+\frac{N}{2},-\frac{m_Q^2}{4k_\perp^2}\right)\\
&+\frac{4}{(1+N)^2}{}_2F_1\left(1,\frac{1+N}{2},\frac{3+N}{2},-\frac{m_Q^2}{k_\perp^2}\right)\\
&-\frac{2^{1-N}}{(1+N)^2}{}_2F_1\left(1,\frac{1+N}{2},\frac{3+N}{2},-\frac{m_Q^2}{4k_\perp^2}\right)\\
&+\frac{2m_Q^2}{k_\perp^2(3+N)}{}_2F_1\left(1,\frac{3+N}{2},\frac{5+N}{2},-\frac{m_Q^2}{k_\perp^2}\right)\\
&-\frac{2^{-2-N}m_Q^2}{k_\perp^2(3+N)}{}_2F_1\left(1,\frac{3+N}{2},\frac{5+N}{2},-\frac{m_Q^2}{4k_\perp^2}\right)\\
&-\frac{2^{-1-N}(16+17N+5N^2)}{N(1+N)(2+N)}\log 2+\left(-\frac{2}{N}+\frac{3+N}{2+3N+N^2}\right)\log\left(1+\frac{m_Q^2}{k_\perp^2}\right)\\
&+\frac{2^{-2-N}(16+17N+5N^2)}{N(1+N)(2+N)}\log\left(4+\frac{m_Q^2}{k_\perp^2}\right)\Bigg), \quad \text{for } N>0\,,
\end{aligned}
\tag{5.107}
$$

where, for $N=0$ we have

$$
\begin{aligned}
\mathcal{I}\left(0, g_{\rm ir}\,\middle|\, Q_{\rm uv}\right) = C_F\Bigg(&-\frac{5}{8}-\frac{2k_\perp}{m_Q}\left[\arctan\left(\frac{m_Q}{k_\perp}\right)-\arctan\left(\frac{m_Q}{2k_\perp}\right)\right]+\frac{1}{2}\log\left(1+\frac{m_Q^2}{k_\perp^2}\right)\\
&-\frac{k_\perp^2}{2m_Q^2}\log\left[4\left(1+\frac{m_Q^2}{k_\perp^2}\right)\right]+\frac{1}{8}\log\left[1+\frac{m_Q^2}{4k_\perp^2}\right]+\frac{k_\perp^2}{2m_Q^2}\log\left[4+\frac{m_Q^2}{k_\perp^2}\right]\\
&+\mathrm{Li}_2\left(-\frac{m_Q^2}{k_\perp^2}\right)-\mathrm{Li}_2\left(-\frac{m_Q^2}{4k_\perp^2}\right)\Bigg),
\end{aligned}
\tag{5.108}
$$

which is equal to $-\mathcal{I}\left(0, Q_{\rm ir} \,\middle|\, Q_{\rm uv}\right)$ by unitarity, and thus

$$\begin{aligned}\mathcal{I}\left(N, Q_{\rm ir} \,\middle|\, Q_{\rm uv}\right) = -\mathcal{I}\left(0, g_{\rm ir} \,\middle|\, Q_{\rm uv}\right) + C_F \sum_{j=1}^{N} \binom{N}{j} (-1)^j \Bigg[& \frac{2^{-2-j}}{2+j} \\ & + \frac{2^{-3-j} m_Q^2}{(8+6j+j^2) k_\perp^2} {}_2F_1\left(1, 2+\frac{j}{2}, 3+\frac{j}{2}, -\frac{m_Q^2}{4k_\perp^2}\right) \\ & + \frac{2^{-2-j}(4m_Q^2 - j(k_\perp^2 + 2m_Q^2))}{j(2+j)k_\perp^2} {}_2F_1\left(1, 1+\frac{j}{2}, 2+\frac{j}{2}, -\frac{m_Q^2}{4k_\perp^2}\right) \\ & + \frac{2^{-2-j}(-1+j) m_Q^2}{(3+4j+j^2) k_\perp^2} {}_2F_1\left(1, \frac{3+j}{2}, \frac{5+j}{2}, -\frac{m_Q^2}{4k_\perp^2}\right) \\ & - \frac{2^{-4-j} m_Q^2}{(4+j) k_\perp^2} {}_2F_1\left(1, 2+\frac{j}{2}, 3+\frac{j}{2}, -\frac{m_Q^2}{4k_\perp^2}\right) \\ & - \frac{2^{-2-j}(16+17j+5j^2)}{j(1+j)(2+j)} \log\left(1 + \frac{m_Q^2}{4k_\perp^2}\right) \Bigg] . \end{aligned} \tag{5.109}$$

Part III

QCD and top jets

CHAPTER 6

Unsupervised learning in the metric space of jets

In the first part of this chapter, we demonstrate how the metric space structure induced by the energy mover's distance can be leveraged for the unsupervised tagging of jets according to their progenitor. Namely, we focus on the task of tagging jets initiated by a top quark from a background of jets initiated by light quarks and gluons. By examining the local neighborhood structure of this metric space, we find that the jets of each class populate the landscape in differing densities. This characteristic can be exploited to accurately cluster jets according to their densities through unsupervised clustering algorithms, such as DBSCAN. In the second part of this work, we modify the metric space by reducing the global notion of connectivity down to a local one and, in the process of doing so, modify our distance metric to be that corresponding to geodesics on an underlying graph. We demonstrate how this modification induces regions of both positive and negative values of curvature, which are then exacerbated through a Ricci flow algorithm. Differences in the curvatures averaged over local patches of the new graph metric space then lead to a flow which separates the signal top jets from the background in a fashion that is completely agnostic to any pre-determined jet labels.

6.1 Introduction

In recent years, the adoption of machine-learning (ML) techniques by the scientific community, in general, and the particle physics community, in particular, has become increasingly widespread

[Sch21, PBD22]. On the particle physics front, such tools have found a natural home in the analysis of jets produced by the Large Hadron Collider (LHC) at CERN. Each jet itself is a collimated beam of $\mathcal{O}(10)$ particles, and with each event producing similar orders of such jets, we see such collisions as being one of Nature's many sources of "Big Data." The jet substructure community has fruitfully adapted an enormous number ML and data science techniques to better understand the internal structure of jets [LMN17, Kog19, KMM20, MSS19]. For instance, early on it was realized that jets and their substructure can be naturally visualized and represented in the data structure of an image [CKS15] and subsequently passed through deep convolutional neural networks (DCNNs) for classification tasks [OKM16, KMS17, CK18]. Variational autoencoders (VAEs) have provided a means of compressing the internal features of jets down to a low-dimensional latent representation [DFK20, DPS21, Col21]. Anomaly detection has seen a great deal of development in efforts to uncover any potential signs of physics beyond the Standard Model (SM) lurking in the vast data sets produced by the LHC [FHM22, Kas21, BDF22]. For an extensive list that covers the many applications of ML to particle physics, see [FN21]. Additionally, ML techniques have allowed the community to enhance its understanding of the SM itself—particularly in the identification and analysis of substructure features produced from the decays of heavy particles, such as top quarks. Top-tagging has become an important arena in this domain [ACM17, KPR17, HKP19, MCD20], for a review, see [But19].

However, the vast majority of the aforementioned applications rely on neural networks that are trained with labeled data, that is machines that "learn" through examples where the identity of a particular jet is known, and then using this learned information to accurately classify new instances of unlabeled jets. Accordingly, such studies fall under the realm of supervised learning. While the performance of such neural networks is impressive, as physicists, we would ideally like to understand and interpret what ML techniques are learning—particularly in terms of quantities for which there exists a physical underpinning and from which one can build intuition, as in the spirit of [FTW21, BHM22]. In doing so, further applications can dispense of the need for labeled data and learn the physical features impressed upon the data itself. Doing so would allow the community to graduate to the realm of unsupervised learning.

An important step in this direction has been in the unsupervised clustering of jets initiated by

heavy resonances such as the W and Z bosons, as well as top quarks through use of an attention mechanism on jet images [MC21]. In addition, concepts from optimal transport (OT) have found a natural home in the realm of jet substructure. The so-called Energy Mover's Distance (EMD) [KMT19c, KMT20] provides a natural metric on the space of jets, allowing one to associate a "cost" for redistributing the substructure of one jet into that of another. A rich geometry results from such a construction, where various familiar event/jet shape observables can be understood as arising from projections onto submanifolds in this abstract space [KMT20]. The richness and utility of this construction has led to the investigation of various extensions, modifications, and alternatives [CCC20, CCC22]. Significant work has been carried out in order to best visualize this space by embedding it into a lower dimensional space in such a way to preserve the salient features of the true manifold [PHO22].

The primary goal of this work is to demonstrate how one can leverage this geometry in order to cluster QCD and top jets in a way that makes no use of labels for jets, i.e. in a completely unsupervised fashion. This is to say, how one can create a top-tagger by accessing only the geometric information afforded by the EMD itself.

This chapter is organized as follows. In Sec. 6.2 we show how QCD and top jets populate their respective metric spaces induced by the EMD. Quantities of interest will be the distribution of distances between jets, the effective dimensionalities of the respective subspaces the jets populate, as well as notions of nearest-neighbor distances. These considerations reveal distinct differences in the geometry of QCD and top jets and lay the groundwork for subsequent sections which work to utilize these differences for the purpose of top-tagging.

In Sec. 6.3 we demonstrate how the preceding geometric structure can be used to cluster QCD from top jets via the unsupervised density-based clustering algorithm, DBSCAN. This provides us with an effective means of tagging top jets from QCD jets, purely through the use of the underlying geometry of the data.

In Sec. 6.4 we demonstrate how QCD and top jets can be separated geometrically through the process of Ricci flow. Ricci flow relies on exacerbating local curvatures in order to separate distinct community structures that live in a weighted graph, and thus, we begin this section by explicitly

laying out the mapping of the EMD space to a modified graph metric space.

Finally, in Sec. 7.5 we conclude and provide an outlook for other interesting jet-tagging tasks that could leverage the underlying data geometry to do so in an unsupervised fashion. As top-tagging has by now become a somewhat canonical task and test bed for ML applications, we compare our work to various other ML-based top-taggers in the literature. While our methods achieve competitive accuracies ($\gtrsim 90\%$), they are not state-of-the-art in this particular sense. However, we argue that our methods clearly distinguish themselves as the only ones that are (1) completely unsupervised, (2) rely on $\mathcal{O}(1)$ input parameters rather than the standard, which is often more than $\mathcal{O}(10^5)$, and (3) have a clear physical interpretation.

6.2 EMD landscapes of QCD and top jets

The physical intuition underpinning this work is that jets initiated by light quarks and gluons versus those initiated by top quarks have fundamentally different internal substructures—this is a well-known fact in the jet substructure community. The statistical hypothesis from which this work begins is that the labels for the jet progenitors imprint themselves on their respective substructures, and such labels can be inferred or characterized by carefully analyzing ensembles of such substructures. What is to be demonstrated in this work is that such labels reveal themselves to high accuracy through the geometry of the data itself. This section will serve to lay the groundwork for the rest of the chapter, as well as convince the reader why such a thing seems plausible. Subsequent sections will provide the quantitative machinery required to convincingly demonstrate this fact.

To begin, we must define the underlying structure of the data from which the aforementioned geometry emerges. Namely, we represent our jets as an unordered set of massless particles, where the i^{th} particle is specified by (p_{Ti}, η_i, ϕ_i) coordinates. Normalizing each p_{Ti} by the total jet p_T^{jet}, $z_i \equiv p_{Ti}/p_T^{\text{jet}}$, the jet is then represented as

$$\mathcal{J}(\eta, \phi) = \sum_{i \in \text{jet}} z_i \, \delta(\eta - \eta_i) \, \delta(\phi - \phi_i) \ . \tag{6.1}$$

This data-type is known as a point-cloud. Given this form, we see (from a statistical viewpoint) that we can treat any given jet as a discrete probability distribution with support on a subset of

the 2D Euclidean plane defined by calorimeter cells on which its constituents are deposited after a collision. It is important to note here that a set of pre-processing steps are required to cast a jet in this format [CKS15, OKM16]. These steps consist of rotations, reflections, and translations to get the jet into a standardized form with common origin in $(\eta\text{-}\phi)$ space defined at the jet's center.

To begin, let us define the indexing set for our data set

$$I \equiv \{0, 1, \ldots, N\} \,. \tag{6.2}$$

In what follows we will consider a data set consisting of QCD and top jets, two thousand of each, all in the point-cloud format of Eq. (6.1), and we will denote this data set by D:

$$D = \left\{ \mathcal{J}_{f,i} \left(\eta, \phi\right) \,\middle|\, f \in \{\text{QCD, top}\} \,,\, i \in I \right\} \,, \tag{6.3}$$

where the first index f runs over the jet flavor labels and i over the indexing set of each ensemble. Our ensembles and labels come from the public top-tagging data set of [But19]. The jets therein are generated through simulation with Pythia8 [Bie22] for collisions at $\sqrt{s} = 14$ TeV, such as those at the LHC. Detector effects are modeled by Delphes [FDD14] paired with use of the ATLAS card. Jets are clustered with the anti-k_T algorithm [CSS08] via FastJet [CSS12], with a jet radius of $R = 0.8$. Each is the leading jet of their respective event. Finally, the transverse momentum range of the data set is $p_T^{\text{jet}} \in [550, 650]$ GeV.

For this section, as well as the remainder of the work, it is important to clarify our use of the jet flavor index. For this present section, we will make ample use of these flavor labels as a means to gain intuition and understanding for the qualitative differences between QCD and top jets as exemplified by use of the EMD. In later sections, this flavor index will only be utilized as the "truth label" to assess the accuracy of each unsupervised classification task performed, as the goal of this work is to tag top jets in a completely unsupervised manner, that is, without any reference to labels affiliated with the data.

With our data set D in hand, we may utilize the energy flow package [KMT19a] to compute EMDs between all pairs of jets, that is we produce a matrix of distances whose $(i\text{-}j)^{\text{th}}$ element will be given the shorthand

$$\text{EMD}_{ij} \equiv \text{EMD}\left(\mathcal{J}_i, \mathcal{J}_j\right) \,. \tag{6.4}$$

As D contains jets with transverse momenta $p_T^{\text{jet}} \in [550, 650]$ GeV, and this fact is controlled for at the level of the point-cloud representation of Eq. (6.1) and results in the EMD taking on a dimensionless value less than unity. We may interpret this distance then as $\text{EMD} \sim Q/p_T^{\text{jet}}$ where Q is the "traditional" EMD with units of GeV and p_T^{jet} is in the range specified above. As the dimensionful EMD has the interpretation of the amount of work (angular) work required to redistribute the energy of one jet into another, the dimensionless/normalized EMD then can be interpreted as the scale of angular resolution that must be traversed in morphing one jet's point-cloud into another.

Our data set D together with the distances between all elements, as computed through the EMD, furnishes us with a metric space $M = (D, \text{EMD})$. The simplest inspection we may perform in the metric space involves visualizing the distribution of EMDs. In doing so, we have three distinct distributions to consider: first, the distances purely between QCD jets (we will refer to this as the "QCD-QCD" distribution), second, those purely between top jets ("top-top") and third, the EMDs between QCD and top jets ("QCD-top"). These distributions are displayed in Fig. 6.1. This basic figure contains a wealth of information. To begin, looking just at the first moments of the QCD-QCD and top-top distributions, we see that the bulk of QCD jets situate themselves far closer to one another than do top jets—this suggests that the QCD subspace of our EMD metric space is populated at a higher density than is the top subspace. The next striking feature is that the QCD-top distribution nearly overlays the top-top. This is to say that there exists similar spacing between top jets amongst themselves as there does QCD and top jets. Pairing this with the fact that the QCD-QCD distribution is shifted towards lower EMD values relative to the top-top and QCD-top distributions, a significant separation between the QCD jets and the top jets is suggested. Thus Fig. (6.1) suggests two qualitative features of the data set: (1) that QCD jets populate a region of higher density than do top jets and (2) these two regions are widely displaced from one another.

Let us further investigate the differing densities in which QCD and top jets populate M. To do so, we will define an important class of sets. Sets of this class will be defined for each $\mathcal{J}_i \in D$,

$$\text{EMD}_i \equiv \left\{ \text{EMD}_{ij} \,\middle|\, j \in I \,,\, j \neq i \right\} \,, \tag{6.5}$$

which is simply the set of EMDs between the jet $\mathcal{J}_i$ and all other jets in D. We use the notation

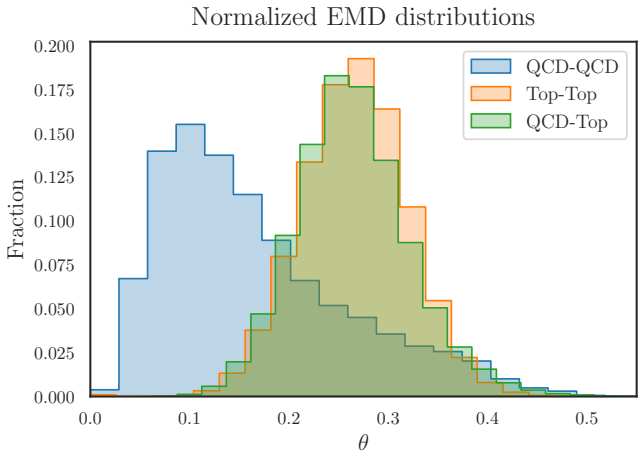

Figure 6.1: The three distinct distributions of EMDs affiliated with our data set of QCD and top jets. The QCD-QCD distribution (blue), the top-top distribution (orange), and the QCD-top distribution (green).

EMD_i to elicit thoughts of "taking the i^{th} row" of the matrix EMD_{ij}, but it is important to note that while EMD_i contains the elements of the i^{th} row of EMD_{ij}, it is just a set and therefore has no notion of "location" for any particular element, as would be captured by the element's column number within the row. We opt for defining this class of sets so that we can order their elements in what follows.

Next, we will define the κ-EMD for each jet $\mathcal{J}_i \in D$ to be the K^{th} order statistic[1] of the set EMD_i:

$$\kappa_i = \mathrm{EMD}_{i(K)}\,, \tag{6.6}$$

which is simply the EMD between $\mathcal{J}_i$ and its K^{th} nearest-neighbor. We choose this notation to highlight the fact that κ_i is a particular matrix element. Thus, given the data set D, one can construct distributions in κ for each jet type as well as choice in K. Doing so creates distributions in κ parameterized by the nearest-neighbor number K and the jet flavor label f, which we denote through standard notation $p_f\left(\kappa; K\right)$. An interesting quantity to look at will be the first moments of

[1]For a set of objects in which there exists a natural ordering (as is the case of EMDs as they are non-negative real numbers) the k^{th} order statistic is the k^{th} smallest element, according to the natural ordering, and is denoted $X_{(k)}$. For example, given a set $X = \{X_1, \ldots, X_n\}$, $X_{(1)} = \min X$ and $X_{(n)} = \max X$.

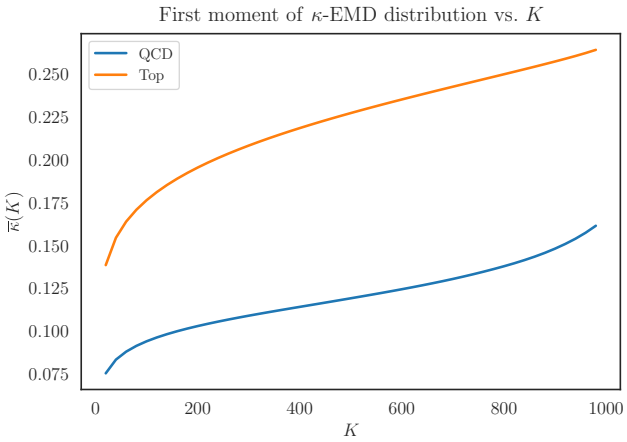

Figure 6.2: First moments of the κ-EMD distributions for QCD (blue) and top (orange) jets, plotted as functions of the nearest-neighbor number K.

such distributions as functions of their conditional variables

$$\overline{\kappa}_f(K) \equiv \int d\kappa\, \kappa\, p_f\left(\kappa; K\right) \ . \tag{6.7}$$

A plot containing curves for $\overline{\kappa}_{\text{QCD}}(K)$ and $\overline{\kappa}_{\text{top}}(K)$ is displayed in Fig. 6.2. A striking feature is the clear ordering between the two flavor classes. For every value of K, the average distance between a QCD jet and its K^{th} nearest-neighbor is roughly half that of tops. This corroborates the intuition gleaned from Fig. 6.1, namely that the QCD jets are more tightly packed, and thus occupy a region of higher density in M than do top jets. We will return to this plot when it comes to interpreting the optimal values of input parameters for the DBSCAN algorithm in Sec. 6.3.

The set of features to investigate will be the correlation dimensions [GP83, K02, Cam03] for the QCD and top subspaces, as done in the case of W-jets in [KMT19c] as well as both quark and gluon jets in [KKT22]. The correlation dimension is computed as a function of resolution scale θ, according to [KMT19c]

$$d_{\text{corr}}\left(\theta\right) = \frac{\partial}{\partial \log \theta} \log \sum_{1 \leq i < j \leq N} \Theta\left(\text{EMD}_{ij} < \theta\right) \ , \tag{6.8}$$

where $\Theta(\cdot)$ denotes the indicator function for the constraint listed as its argument. From a geometric standpoint, smaller θ resolves local structure while larger θ resolves more global features of the metric space. Thus from the physics perspective, the local structure of M reflects IR information

of the jet, while the global structure captures the UV. We display curves of d_{corr} for the QCD and top subspaces in Fig. 6.3.

Let us now analyze Fig. 6.3 in detail. We see that the correlation dimension for QCD jets takes the form of a power-law, while that of top jets exhibits a more non-trivial θ-dependence, with the key being an exponential decay in the $\theta \gtrsim 0.3$ region. Both of these observations can be understood as manifestations of the characteristic energy scales for QCD and top jets.

For QCD jets initiated by light quarks a gluons, the characteristic energy scale is $\Lambda_{\text{QCD}} \sim \mathcal{O}(100)$ MeV, which compared to $p_T^{\text{jet}} \sim \mathcal{O}(100)$ GeV, is effectively zero. This is to say that the characteristic angular scale for QCD jets is $\theta_{\text{QCD}} \sim 0$, or in other words, that QCD jets are approximately scale-free. This lack of definitive angular scale leads to the power-law behavior $d_{\text{corr,QCD}}(\theta) \sim \theta^{-n}$ for some $n > 0$.

This is in sharp contrast to the top jets, which are distinguished by their characteristic energy scale $m_{\text{top}} \sim 173$ GeV. This leads to an angular scale of around $\theta_{\text{cd}} \sim m_{\text{top}}/p_T^{\text{jet}} \in [0.27, 0.31]$. We denote this quantity by θ_{cd}, to allude to the fact that it defines the approximate angular extent of the top's characteristic decay $t \to q\bar{q}'b$. Looking back to Fig. 6.3 we see that angular scales $\theta > \theta_{\text{cd}}$ are too large to resolve any of the substructure features that differentiate top jets from one another, i.e. separate them in the EMD metric space. This gives the top subspace a vanishing correlation dimension. Once we probe the region $\theta \sim \theta_{\text{cd}}$, characteristic substructure is resolved, and the complexity of the top jet's population in M becomes manifests itself in a rapid growth in $d_{\text{corr,top}}(\theta)$. In the opposite limit, i.e. $\theta < \theta_{\text{cd}}$ we recover an approximate power-law dependence which can be understood as follows. In going far enough below this resolution scale, the top mass is effectively integrated out and we are left with a metric space population pattern characteristic of a scale-free theory like QCD. Physically what is happening is that in probing the metric space at small angular resolution, we are probing neighborhoods of jets who differ only through subtle variations in radiation patterns, generated through QCD evolution, falling collinear to the three (on average) hard prongs marking the characteristic decay pattern of top jets.

The most significant fact regarding the correlation dimensions of Fig. 6.3 is the striking disparity in their numerical sizes amongst the two jet flavors. At any reasonable angular resolution scale

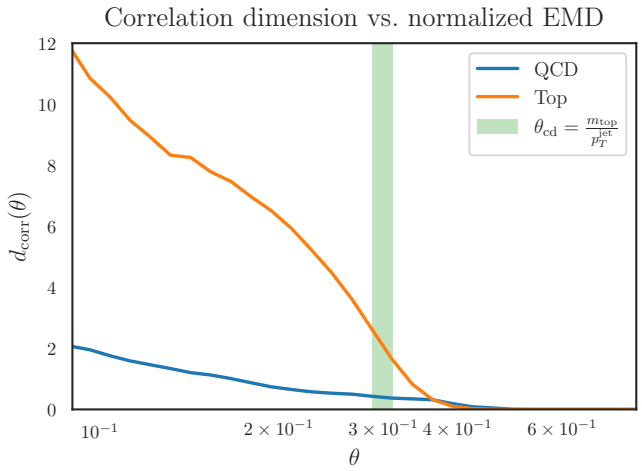

Figure 6.3: Correlation dimensions for the QCD (blue) and top (orange) subsets of the metric space plotted as a function of the EMD resolution variable θ. The green band corresponds to the range of values taken on by the top's characteristic decay angle θ_{cd}, whose finite width is due to $p_T^{\mathrm{jet}} \in [550, 650]$ GeV. This marks the transition region for the resolution of top jets' internal substructures at the level of the geometry of the data.

$\theta < R$, where R is the jet radius, QCD jets populate a region of vastly lower dimensionality than do top jets. This provides yet another indication of the high-density population pattern of QCD jets relative to tops in EMD space. For consider embedding both subspaces into some larger space that contains both. At every resolution scale, the QCDs are confined to a low-dimensional subspace while the tops are free to diffuse across a larger subvolume.

In total, these three basic considerations regarding the subspaces populated by QCD and top jets indicate that these particular flavor labels imprint themselves in a non-trivial way on the geometry of their resulting metric space. The goal of the following sections is to leverage these geometric features to infer such flavor labels.

6.3 Density-based clustering

6.3.1 DBSCAN background

In this section, we will give a conceptual overview of the DBSCAN algorithm, as implemented in Scikit Learn [PVG11]. Our treatment will closely follow that laid out in the original work [EKS96], however instead of describing the general workings, we will tailor things to describe our specific problem at hand. For a more recent treatise, see [SSE17].

What makes DBSCAN particularly well-suited to the problem at hand lies in its reliance on two simple parameters that work to quantitatively define a density threshold for elements living in a metric space. Or, to state things another and equivalent way, what makes our data set so amendable to density-based clustering is the presence of a rather sharp density demarcation in the EMD landscape. As alluded to in the previous sections, when studied individually, sets of QCD- and top-initiated jets populated EMD space with strikingly different densities. While this fact does not guarantee separability in their combined set, it certainly makes such a feat seem promising. Here, we will show that QCD and top jets are indeed separable according to their densities and the requisite density threshold that separates the domains may be extracted from the data in a label-free fashion.

First, let us define several concepts that DBSCAN makes use of. We will make reference to our data set of jets as D. The first concept will be that of an ϵ-ball about a given jet, as considered in [KMT20],

$$B_\epsilon\left(\mathcal{J}_i\right) = \left\{\mathcal{J}_j \in D \,\middle|\, \mathrm{EMD}_{ij} \leq \epsilon\right\} \,, \tag{6.9}$$

that is, the set of jets within an EMD of ϵ from the given jet $\mathcal{J}_i$. The next point has to do with two qualitatively different types of points belonging to a cluster, namely what are referred to as *core* and *border* points [EKS96]. A core point is defined as one whose ϵ-ball has a cardinality bounded below by some value, μ. This is to say that the jet $\mathcal{J}_i$ is a core point if

$$\left|B_\epsilon\left(\mathcal{J}_i\right)\right| \geq \mu \,. \tag{6.10}$$

This bound is fixed in order to contrast core points from border points, where one can easily picture a point lying on the border of a cluster to have fewer jets surrounding it, that is $\left|B_\epsilon\left(\mathcal{J}_i\right)\right| < \mu$. Next,

the jet $\mathcal{J}_B$ is said to be *directly density-reachable*, with respect to (ϵ, μ), by the jet $\mathcal{J}_A$ if $\mathcal{J}_A$ is a core point and $\mathcal{J}_B \in B_\epsilon(\mathcal{J}_A)$. If $\mathcal{J}_B$ then fails to meet the core point condition itself, it is a border point. Thus, we see that the basic notions of core and border points are defined with respect to the parameter pair (ϵ, μ), and therefore in what follows, we will consider all further notions regarding the connectivity of clusters in our metric space to be defined with respect to this pair.

Jets $\mathcal{J}_A$ and $\mathcal{J}_B$ are said to be *density-reachable* if there exists a sequence of jets $(\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_K)$ such that $\mathcal{J}_1 = \mathcal{J}_A$, $\mathcal{J}_K = \mathcal{J}_B$ and $\mathcal{J}_{i+1}$ is directly density-reachable from $\mathcal{J}_i$. Jets $\mathcal{J}_A$ and $\mathcal{J}_B$ are then said to be *density-connected* if there exists a jet $\mathcal{J}_C$ from which both $\mathcal{J}_A$ and $\mathcal{J}_B$ are density-reachable.

DBSCAN then defines a *cluster*, $C \subset D$ according to the following two conditions: (1) for all $\mathcal{J}_A$ and $\mathcal{J}_B$, if $\mathcal{J}_A \in C$ and $\mathcal{J}_B$ is density-reachable from $\mathcal{J}_A$, then $\mathcal{J}_B \in C$. (2) for all $\mathcal{J}_A, \mathcal{J}_B \in C$, $\mathcal{J}_A$ is density-connected to $\mathcal{J}_B$.

Lastly, in contrast to a cluster, DBSCAN defines the *noise*, $\text{noise} \subset D$, to be that which belongs to no cluster, for example, suppose $C_1, \ldots, C_K$ are clusters of D, then

$$\text{noise} = \left\{ \mathcal{J} \in D \,\middle|\, \mathcal{J} \notin C_i\,,\ \forall i \in \{1, 2, \ldots, K\} \right\}\,. \tag{6.11}$$

To restate intuitively, the noise set is the set of jets where none of its interior points satisfy the core point condition of Eq. (6.10), which is to say that these points fall below the density threshold established by the parameter pair (ϵ, μ).

6.3.2 Unsupervised extraction of input parameters

In the previous section, we outlined how DBSCAN defines points based on neighborhoods, and then forms clusters based on the connectivity of such neighborhoods. All definitions were ultimately made with respect to the parameter pair (ϵ, μ). This pair forms the basic user-input to the DBSCAN algorithm—using the notation of Scikit-learn [PVG11], $(\epsilon, \mu) \to (\texttt{eps}, \texttt{min\_samples})$. Now, as the goal of this work is to perform top-tagging in an unsupervised way, we need a method to determine $(\epsilon, \mu) \to (\texttt{eps}, \texttt{min\_samples})$ without the use of any labels on our data set. In what follows, we describe a particular procedure for doing so.

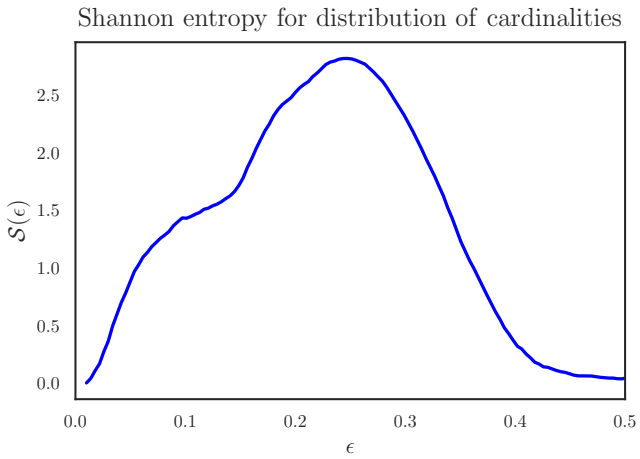

Figure 6.4: The Shannon entropy of the cardinality distribution as a function of ball radius ϵ. The location of the local minimum at $\epsilon \sim 0.1$ determines the ideal value of the `eps` parameter to be used by DBSCAN.

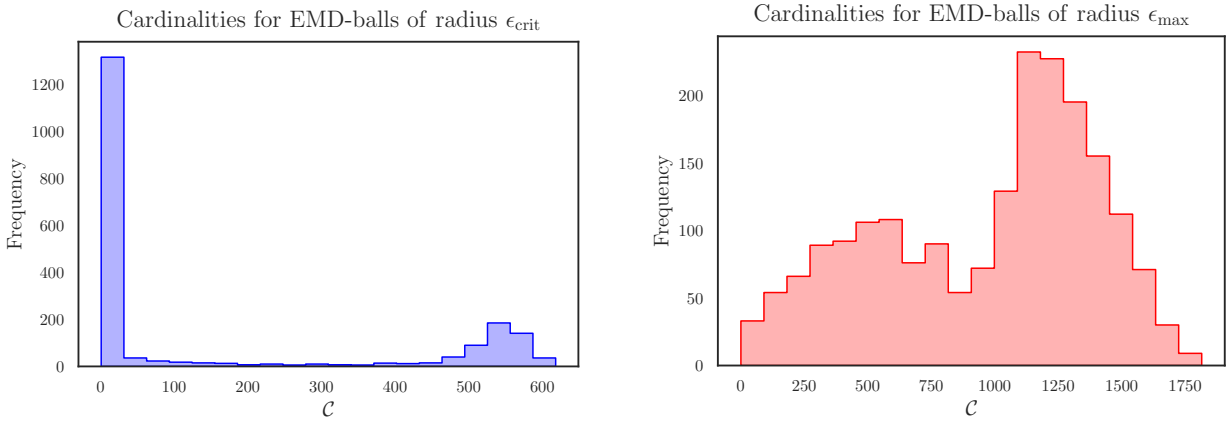

Figure 6.5: Cardinality distributions corresponding to the two extremum points. (Left) Distribution corresponding to $\epsilon_{\text{crit}} \sim 0.1$, where a well-separated bimodal structure is clear. (Right) Distribution corresponding to $\epsilon_{\text{max}} \sim 0.5$, where the bimodal structure is smeared to a large degree.

Our procedure is rather simple. First, we consider the entire unlabeled data set consisting of an equal mixture of QCD and top jets. We construct ϵ-balls about each jet and then compute the resulting cardinality of each ball. The resulting cardinality then ends up being dependent on both ϵ and the particular jet $\mathcal{J}$:

$$\mathcal{C}\left(\epsilon, \mathcal{J}\right) \equiv \left|B_\epsilon\left(\mathcal{J}\right)\right| \, . \tag{6.12}$$

Next, we collect the resulting cardinalities $\forall \mathcal{J} \in D$ and bin them into a histogram, normalizing by the total number of jets $N + 1 = |D|$. We thus consider the resulting histogram to be a discrete distribution in the random variable $\mathcal{C}$, whose stochasticity is derived from that of the underlying set of jets forming D, and is conditional on the variable ϵ. This is to say, that for each value of ϵ, we have a probability distribution $p\left(\mathcal{C}; \epsilon\right)$, where, despite using "continuum" notation, we keep in mind that $\mathcal{C}$ is indeed discrete. To state in other words, the EMD scale ϵ defines a one-parameter family of cardinality distributions $p\left(\mathcal{C}; \epsilon\right)$.

Lastly, we consider the one-parameter family of Shannon entropies [CT06] that results from our one-parameter family of cardinality distributions

$$\mathcal{S}(\epsilon) = -\sum_{\mathcal{C}} p\left(\mathcal{C}; \epsilon\right) \log p\left(\mathcal{C}; \epsilon\right) \, . \tag{6.13}$$

The Shannon entropy quantifies the number of bits (if one uses log base 2) required to specify the distribution. From a more qualitative standpoint, the entropy is a proxy for how much "uncertainty" lies in the random variable $\mathcal{C}$. This being the case, we can immediately point understand two special limits. First, we have that

$$\lim_{\epsilon \to R} \mathcal{S}(\epsilon) = 0 \, , \tag{6.14}$$

since $R \gtrsim \max\left\{\mathrm{EMD}_{ij} \,\middle|\, i, j \in \{0, 1, \ldots, N\}\right\}$ is the maximum angular distance scale occurring in D, for in this scenario all balls have a cardinality n, so the entirety of the distribution's mass lies in the bin containing n. So, since $\mathcal{S} \to 0$ at both endpoints and $\mathcal{S} \geq 0$, the mean value theorem tells us $\mathcal{S}$ has a global maximum for some value of ϵ. Now, the scale at which $\mathcal{S}$ reaches its global maximum is the scale at which $p\left(\mathcal{C}; \epsilon\right)$ is the closest it comes to a uniform distribution, which is to say the furthest it comes to any sense of bimodality.

It turns that the closest approach to bimodality in $p\left(\mathcal{C};\epsilon\right)$ occurs at the ϵ-value for which $\mathcal{S}$ achieves a saddle-point. As ϵ increases above zero, bin content from redistributing itself from 1 to populate higher values of $\mathcal{C}$—doing so leads to a sharp increase in $\mathcal{S}$. As this continues, bins of higher and higher cardinality are populated, which extends the support set of $p\left(\mathcal{C};\epsilon\right)$. Eventually, this rate slows down and reaches a local extremum, as depicted in Fig. 6.5. At this point, our distribution is the approximately bimodal, with a large peak at $\mathcal{C}=1$ and a localized Gaussian hump near $\mathcal{C}\approx 550$. We refer to this critical value as ϵ_{crit}. As we move into the region $\epsilon \gtrsim \epsilon_{\text{crit}}$, more and more bin content from 1 bleeds over into the "valley" separating the two peaks, hence lessening their respective levels of localization. This occurs until we reach ϵ_{max} at which the global max of $\mathcal{S}$ occurs, where the dual-peak structure is most obscured. After this point, the maximally-spread distribution starts localizing/collapsing to the single peak at n.

We find that $\epsilon_{\text{crit}}\approx 0.1$ and we can understand the emergence of this scale as follows. Consider the EMD distributions for QCD and top jets separately, as displayed in Fig. 6.1. We find that, in this plot, $Q/p_T \sim \epsilon_{\text{crit}}$ marks the lower-tail region of the top distribution, this is to say that there only exist QCD jets with inter-jet distances below ϵ_{crit}. Once we reach this threshold, top jets start coming into the mix. Therefore, ϵ_{crit} provides a natural ball-radius to assess QCD densities. Displayed in the (b) panel of Fig. 6.5 is the distribution $p\left(\mathcal{C};\epsilon_{\text{crit}}\right)$, where the first bin is expected to be largely populated by top jets, and higher bins, particularly those for $\mathcal{C}>300$, are expected to be QCD jets, for these jets must be very dense. It is worth stressing again that while our hypothesis of the labels corresponding to bin contents is inspired by the analysis of previous sections, this entire procedure makes no use of labels. On general grounds, the saddle-point of $\mathcal{S}(\epsilon)$ is a natural scale to emerge, and inspection of $p\left(\mathcal{C};\epsilon_{\text{crit}}\right)$ reveals a rather striking hint at the presence of two clusters. The vast valley between peaks makes the choice of $\mathcal{C}$ cut quite natural as well. Let us call the midpoint of the valley separating the two peaks $\mathcal{C}=\mu_{\text{crit}}$ just to pair with ϵ_{crit}.

Thus, this simple procedure of examining critical EMD value for the one-parameter family of Shannon entropies resulting from the one-parameter family of cardinality distributions provides us with natural input parameters for DBSCAN in a completely unsupervised way. Thus we take

$$\left(\texttt{eps},\,\texttt{min\_samples}\right)=\left(\epsilon_{\text{crit}},\mu_{\text{crit}}\right), \tag{6.15}$$

so that DBSCAN defines a core point of a cluster to be

$$|B_{\epsilon_{\text{crit}}}(\mathcal{J}_{\text{core}})| \geq \mu_{\text{crit}} \,. \tag{6.16}$$

Interestingly, based on our knowledge from the previous section, we see that by-and-large only QCD jets will meet this criterion, and thus we can expect DBSCAN form a single cluster of QCD jets and classify the top jets as noise. This is natural for our binary data set, as only a single density threshold need exist to partition our data.[2]

6.3.3 Top-tagging results from density-based clustering

Insertion of the EMD matrix for our data set and the application of the parameter values as determined in Eq. (6.15) to DBSCAN is the full extent of the initialization required for our clustering task. With these specifications, DBSCAN outputs the identification of one set identified as a cluster, which recalling our discussion in Sec. 6.3.1, means that one subset of our data meets the criteria of density-reachability and density-connectivity according to Eq. (6.16), while the other subset does not meet these constraints and is thus identified as noise. Our analysis in Sec. 6.2 then clearly identifies QCD jets as those forming the cluster with tops making up the noise. The accuracy of this clustering is then assessed through the use of the F_1 score, whose value is determined via

$$F_1 = \frac{2\text{tp}}{2\text{tp} + \text{fp} + \text{fn}} \,, \tag{6.17}$$

where tp denotes a true positive, fp a false positive, and fn a false negative. Such is a standard metric for the assessment of the accuracies of unsupervised clustering tasks. The accuracy we obtain is

$$F_1\Big|_{\text{EMD+DBSCAN}} = 0.9003 \,. \tag{6.18}$$

We can visualize the efficacy of the clustering task through the embedding of the EMD metric space in a two-dimensional space obtained through use of the dimensionality-reduction algorithm

[2] If we had additional jet types, say W- or Z-jets, occupying the metric space in differing densities, our method could be applied using in successive steps, still using DBSCAN. Alternatively, one could in principle cluster the varying densities using a more sophisticated clustering algorithm, such as HDBSCAN [MHA17]. We leave such investigations to future work.

UMAP [SMG21]. In our initialization of UMAP, we specify a denseMAP [NBC20] value 2.0 in order to preserve some of the local density profiles exhibited in the EMD space. Fig. 6.6 displays the UMAP embeddings of our data set labeled according to both their true labels, as well as their cluster assignments determined by DBSCAN.

We see that the vast majority of QCD jets are concentrated in a dense pocket in the lower right corner of the embedding, while only a couple handfuls or so are dispersed over the wide and sparsely-populated region dominated by the tops. Again, in light of our discussion in Sec. 6.3.1 and core/cluster point criterion defined by Eq. (6.16), we can understand why DBSCAN cannot extend the QCD cluster too far into the top/noise landscape.[3] Looking back to Fig. 6.1, we see that relative shift towards higher EMD values of the QCD-Top distribution relative to the QCD-QCD distribution is what allows for the density criteria to be satisfied by rather pure sample of QCD jets with minimal top contamination. This shift is then visualized in Fig. 6.1 by the separation of each class's effective center-of-mass in UMAP space.

We remark that the physical origins of such a shift can be traced back to the presence of the additional scale (m_{top}) characterizing the top jets which is absent in the QCD jet sample. That scale not only imprints itself upon the Top-Top EMD distribution, but more importantly does do almost equivalently on the QCD-Top distribution as well. Thus, one may reasonably inquire about whether such a feature is general to jets stemming from the decay of heavy-resonances relative to approximately scale-free QCD jets. We leave such interesting investigations to future work.

6.4 Ricci flow on graphs

6.4.1 Overview and graph-theoretic preliminaries

In this section, we provide an alternative method for the unsupervised clustering of QCD and top jets that, like the previous section, utilizes only the geometric information induced by the EMD. We use the algorithm developed in [NLL19] for the unsupervised clustering of subgraphs

[3]Note the few blue points deep in the interior of the top region in the bottom plot of Fig. 6.6 are due the minor distance fluctuations/distortions inherent to the highly-nonlinear embedding function defined by UMAP.

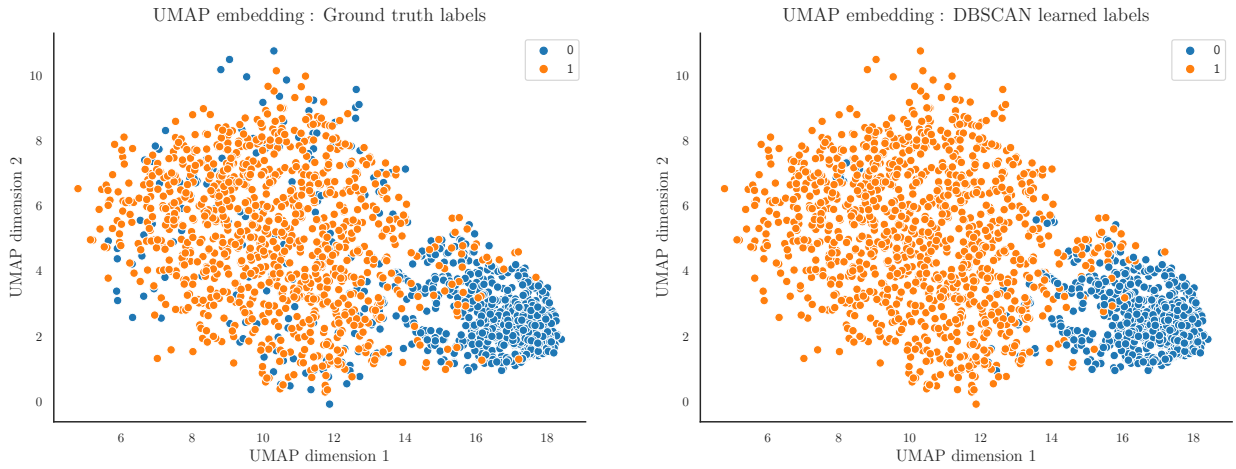

Figure 6.6: UMAP embedding of the data set colored according to their ground truth labels (left) and the cluster labels assigned by DBSCAN (right) : QCD (blue), top (orange).

(or "communities") within the graph defining the data set. The algorithm does this by computing Ricci curvatures between nodes in the graph and then iteratively updating the weights between nodes according to Ricci flow. Loosely speaking, Ricci flow works to compress and stretch regions of positive and negative curvature, respectively. The intuition is then that clusters within the graph, which are connected to one another through regions of negative curvature, can be stretched out and separated as distinct communities. We will outline key technical details behind the algorithm in what follows.

First we must get our data set into a form that is amenable for this Ricci flow process. Doing so first requires a re-interpretation of our metric space $M = (D, \text{EMD})$ as a weighted graph G [Die17]. This can be done in a very straightforward fashion by mapping M to a fully-connected graph G where each jet is represented by a node, with each node having edges connecting it to every other node, and then assigning edge weights given by the EMD distance between jets. The problem with such a procedure, though, is that it results in a fully-connected graph, but the Ricci curvatures for fully-connected graphs are uniformly positive—thus there will exist no regions of negative curvature for the Ricci flow algorithm to dilate and reveal any community structure among subgraphs. This is to say that we won't be able to resolve any clusters of jet type through Ricci flow following this procedure.

However, there exists a simple way to induce regions of negative curvature at the graph level which relies on intuition gained in Secs. 6.2 and 6.3. Namely, we can make use of the κ-EMD defined in Eq. (6.6) to reduce the edge set of a given vertex to define connections to a jet's K-nearest neighbors only, as opposed to the entirety of the data set D. Hence, we can consider each vertex to connect itself to a local patch of jets of a fixed cardinality, analogous the the core-point condition of DBSCAN, Eq. (6.10), but substituting the ball-radius ϵ for κ, and μ for K, we will consider the analogue of a ball in a metric space for an edge set in a graph. Schematically,

$$|B_\epsilon(\mathcal{J})| \geq \mu \longleftrightarrow |B_\kappa(v_i)| = K \,. \tag{6.19}$$

The global connectivity of the graph is then induced via mutual overlap between neighboring patches about vertices. The remainder of this section will be devoted to building up the mathematical machinery required to make this correspondence precise.

Consider our collection of jets of point-cloud data-type and size $N+1$, as in Eq. (6.3). The EMD furnishes us a metric on elements of D, i.e. EMD_{ij} provides a distance between jets $\mathcal{J}_i$ and $\mathcal{J}_j$. This pairing of the set D with the EMD metric furnishes us with a metric space

$$M = (D, \mathrm{EMD}) \,. \tag{6.20}$$

Let us recast M as a graph. A weighted graph G is a collection of vertices V, edges E, and edge weights ω [Die17], i.e.

$$G = (V, E, \omega) \,. \tag{6.21}$$

As previously mentioned, a fully-connected graph is a graph in which for every vertex $v_i \in V$ there exists an edge $v_i v_j \in E$ connecting v_i to v_j for all $v_j \in V$, and it is this structure that we want to modify. As such we will construct a graph whose vertices are connected K other vertices, where K is not necessarily equal to N. Taking $K = N$ will give the fully-connected case but in what follows, we will take K to be general, ultimately linking it to the K that defines the κ-EMD.

To begin, let us consider each jet in our metric space as a vertex in our graph, ultimately endowing the vertex set with the same labeling scheme as our data set

$$V = \left\{ v_i \,\middle|\, i \in I \right\} \,. \tag{6.22}$$

All that differs is just that we are no longer keeping track of the underlying point-cloud nature of each jet. Such information is used only for the computation of EMDs, which we will take as given and will be carried by the edge weights. At this point, we see trivially that there is a one-to-one correspondence between the sets $D \longleftrightarrow V$.

We will consider the mapping from the metric space M to the graph G to be carried out schematically by the function Γ

$$M \xrightarrow{\Gamma} G\,. \tag{6.23}$$

This relation is schematic, as Γ will really be a pair of functions, one for vertices and the other for edge weights.

The simplest instance of this mapping, which we will denote as Γ_{vertex}, provides the correspondence between jets $\mathcal{J} \in D$ and vertices $v \in V$:

$$\begin{aligned} \Gamma_{\text{vertex}} \colon D &\longrightarrow V \\ \mathcal{J}_i &\longmapsto v_i\,, \end{aligned} \tag{6.24}$$

which is to say that the image of D under Γ_{vertex} is V:

$$V = \Gamma_{\text{vertex}}(D)\,. \tag{6.25}$$

Now, given our set of vertices, we will define the edge set for each vertex. In doing so, we will make use of the EMD-set for each jet, defined by Eq. (6.5), and the κ-EMD of Eq. (6.6)

$$E_K(v_i) \equiv \left\{ v_i v_j \,\middle|\, \text{EMD}_i \leq \kappa_i\,,\; j \in I\,,\; j \neq i \right\}\,, \tag{6.26}$$

where we adorn the set with a subscript K to emphasize that the set inherits dependence on K through the κ-EMD constraint. This definition makes it so that each $v_i \in V$ is connected to only K neighbors. As such, the degree of each vertex, that is, the cardinality of the edge set for each vertex is K: $\deg(v_i) \equiv |E_K(v_i)| = K$. A set which comes as a natural byproduct of $E_K(v_i)$ is neighborhood set of each vertex v_i:

$$\nu_K(v_i) \equiv \left\{ v_j \,\middle|\, v_i v_j \in E_K(v_i) \right\}\,. \tag{6.27}$$

We will make use of this set in the following section where we describe the mechanics of the Ricci flow algorithm.

The total edge set for the K-connected graph is then the union of the edge sets of each vertex

$$E_K = \bigcup_{i=0}^{N} E_K(v_i)\,. \tag{6.28}$$

Once the K-restricted edge set E_K is determined, the weight set is simple to define, as there is naturally a one-to-one correspondence between $\omega \longleftrightarrow E_K$. This correspondence is then induced by what we will label as Γ_{weight}:

$$\Gamma_{\text{weight}}\colon E_K \longrightarrow \omega \tag{6.29}$$

$$v_i v_j \longmapsto \text{EMD}_{ij}\,. \tag{6.30}$$

This is to say that ω is the image of E_K under Γ_{weight}:

$$\omega = \Gamma_{\text{weight}}(E_K)\,. \tag{6.31}$$

With the aforementioned sets in hand, we define the K-connected graph G_K to be the collection

$$G_K = (V, E_K, \omega)\ , \tag{6.32}$$

where again, we highlight the fact that the edge set carries dependence on the choice of nearest-neighbor connections K, which ultimately controls all the interesting connectivity properties of the graph.

6.4.2 Graph Ricci flow background

In this section, we provide an overview of the what goes into the computation performed by the Ricci flow algorithm of [NLL19], following their exposition closely. In the previous section, we described the mapping that takes the metric space M to the graph G_K. We will schematically write the algorithm of [NLL19] to be a mapping $\mathcal{R} : G_K \to G'_K$ where G'_K is essentially G_K, but with its set of edge weights modified. The "flow" aspect of the Ricci flow will consist of iterating $\mathcal{R}$ numerous times, resulting in a family of edge-weight-modified graphs $G_K^{(t)}$, where t can be thought

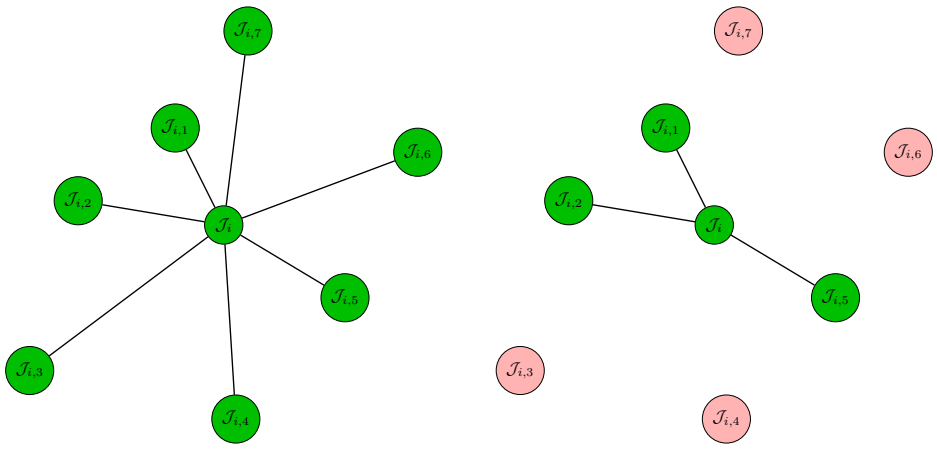

Figure 6.7: Schematic depiction of the modification to the connection structure for a hypothetical jet $\mathcal{J}_i$ to its seven nearest-neighbors after reducing its connectivity to be that to its three. Note that only connections emanating from $\mathcal{J}_i$ are depicted, the connectivity structures of its neighboring points are omitted for clarity.

of as a discretized "time" parameter. The workflow from our original metric space M to the final graph with QCD and top subgraphs separated into clusters proceeds then as

$$M \xrightarrow{\Gamma} G_K^{(0)} \xrightarrow{\mathcal{R}} G_K^{(1)} \xrightarrow{\mathcal{R}} \cdots \xrightarrow{\mathcal{R}} G_K^{(n)} \,. \tag{6.33}$$

To define the Ricci curvature between vertices, we must first define a metric on the graph. We will refer to such an object as $d_G^{(t)}$, which is to say that it will be updated through each iteration of $\mathcal{R}$. This metric will be a function of the edge weights, thus inheriting its time-dependence through that of $\omega^{(t)}$. The metric for a general time t is defined to be

$$d_G^{(t)}\left(v_i, v_j\right) = \min\left\{ \sum_{k=0}^{n-1} \omega^{(t)}\left(v_{A_k} v_{A_{k+1}}\right) \,\middle|\, v_{A_k} v_{A_{k+1}} \in E_K \,,\, v_{A_0} = v_i \,,\, v_{A_n} = v_j \right\}. \tag{6.34}$$

This may be intuitively understood as follows: due to the connectivity of the graph, while not all vertices are directly-connected to one another by a single edge, they are connected through a series of edges which can be thought of as those being traversed in hopping from one vertex to the next. Thus, we consider the set of all possible sequences of edges, sum their corresponding weights, and take the minimum length resulting from such traversals. This being the case, we can naturally relate $d_G^{(0)}$ directly to the EMD, as

$$\omega^{(0)}\left(v_{A_k} v_{A_{k+1}}\right) = \mathrm{EMD}_{A_k A_{k+1}} \,, \tag{6.35}$$

albeit through the non-trivial relation of Eq. (7.15), which is a direct result of the non-trivial connectivity structure of $G_K^{(0)}$.

The process to computing the Ricci curvature between vertices in a graph involves the addition of one more structure, that is a probability measure defined on the graph. This probability measure will depend on the underlying graph metric $d_G^{(t)}$ and will thus inherit the time-dependence therefrom. The authors of [NLL19] define the graph measure to be

$$\mathcal{P}_i^{(t)}(v_j) = \begin{cases} \alpha & \text{if } v_j = v_i \\ \dfrac{1-\alpha}{Z_i} \exp\left(-d_G^{(t)}\left(v_i, v_j\right)^{\beta}\right) & \text{if } v_j \in \nu_K\left(v_i\right) \\ 0 & \text{otherwise} \end{cases} , \tag{6.36}$$

where

$$Z_i \equiv \sum_{v_j \in \nu_K(v_i)} \exp\left(-d_G\left(v_i, v_j\right)^{\beta}\right) , \tag{6.37}$$

is the effective partition function for the i^{th} vertex, providing normalization, and the parameter set (α, β) is taken to be $(1/2, 2)$. The natural metric used for the space of measures is the 1-Wassertein distance W_1, whose formulation falls under the scope of optimal transport [Vil09], just like the EMD [KMT19c, KMT20]. In doing so, we introduce a new metric space $(\mathcal{P}, W_1)$ atop the underlying graph metric space (G_K, d_G). This pair allows us to then compute the Ricci curvature is between vertices v_i and v_j according to

$$R^{(t)}(v_i v_j) = 1 - \frac{W_1\left(\mathcal{P}_i^{(t)}, \mathcal{P}_j^{(t)}\right)}{d_G^{(t)}\left(v_i, v_j\right)} , \tag{6.38}$$

where the curvature is interpreted to be along the shortest path between v_i and v_j. The curvatures then define the update edge weights—the flow then proceeds iteratively as

$$\omega^{(t+1)}\left(v_i v_j\right) = \left(1 - R^{(t)}(v_i v_j)\right) \times d_G^{(t)}\left(v_i, v_j\right) . \tag{6.39}$$

For more details regarding the workings of the algorithm, see [NLL19].

At this point, it is important to comment on the interpretation of the graph resulting from the process of Ricci flow. As stated previously, only the initial graph $G_K^{(0)}$ has direct relation to the

physical EMD, that is has edge weights that can be directly related to angular resolution scale through which jets differ in their physical substructures. Of course, even at the level of $G_K^{(0)}$, while this physical interpretation is valid for each vertex in relation to its K-nearest neighbors, the modified connectivity structure blurs this meaning for distant jets that are connected via paths that require summing over many edges. For subsequent times, the edge weights $\omega^{(t)}$ completely lose their physical interpretation, as they deform according the the will of the flow algorithm. As the the metric space defined by the EMD has a rich manifold structure [KMT20], it would be very interesting to understand any physical interpretation of its evolution through Ricci flow. This of course is far beyond the scope of the current work, which is simply to leverage geometric features in order to tag top jets from QCD. As such we leave such investigations to future work.

6.4.3 Top-tagging results from Ricci flow

From the discussion of the previous section, we see that there is essentially only one parameter having to do with the structure of our data set that we must determine, and that is the nearest-neighbor number K defining the reduced-connectivity structure we impose on our EMD metric space to transform it to that of a graph—see Fig. 6.7 for a visualization of this process. We find that there exists a good deal of freedom in the choice of this value, but values of $\mathcal{O}(10)$ perform the best, which, perhaps not surprisingly, can be understood from our discussion in previous sections.

We note that in Sec. 6.3.2, the determination of ϵ_{crit} was due to it marking an extremum in the Shannon entropy of the cardinality distributions. Fig. 6.5 depicts the cardinality distribution affiliated with the ball radius of ϵ_{crit}. We also know by Fig. 6.1 that this value of ϵ_{crit} marks the peak of the QCD-QCD EMD distribution. Let us combine this information. Fig. 6.1 tells us that the majority of QCD jets are within ϵ_{crit} of one another and Fig. 6.5 corroborates this fact by telling us that the balls around most of these points contain sizeable portions of the QCD subset of the data, since $\mathcal{C} \sim 500$ for these balls. Thus, in choosing $K \sim \mathcal{O}(10)$, while modifying the connectivity drastically (reducing the graph-ball cardinalities by an order of magnitude), the resultant effect on the graph-path metric between QCD jets is expected to be minimal due to the dense-packing of this corner of the metric space. One can conversely anticipate a more drastic change in going

from the EMD distances to the graph-path distances amongst tops, since their EMDs are Gaussian-distributed about $\theta_{\rm cd}$, so the reduction of their connection structure should amplify the number of "$\theta_{\rm cd}$'s" one must hop over to get from one corner of the top landscape to the other. This is all to say, that by reducing the connection structure drastically to $K \sim \mathcal{O}(10)$, we can exacerbate the graph-theoretic differences between QCD and top jets and use Ricci flow to amplify these differences. In the end, we choose $K = 30$ based on these considerations, as $\mathcal{O}(1)$ multiples of this central value all yield similar results. We find that performing fifty iterations of Ricci flow is enough to achieve $\approx 91\%$ accuracy in our top-tagging task.

Interestingly, we find that there are two ways in which we can use the results from the Ricci flow algorithm in order to tag tops jets. The first of which is the most straightforward. This is to simply let Ricci flow run, updating the edge weights/graph-distances between points at each step. This works to effectively separate the portions of the metric space corresponding to QCD and top jets. To visualize this separation, we take the graph-path metric d_G, and embed its values between all elements of the data set into UMAP space, as done in Sec. 6.3.3.

We depict such embeddings in Fig. 6.8, where the top two panels show the embedding after zero (left) and fifty (right) iterations without labels, while the bottom two panels are labeled analogues. We see such a manifest separation of clusters in the UMAP embedding that one can perform a simple cut on the coordinates to ascribe labels. In doing so, we yield an accuracy of 91.04%.

Interestingly, we can achieve similar accuracy through performing a cut on another output of the Ricci flow algorithm. Consider the following object, which can be thought of as analogous to the effective partition function in Eq. (6.37):

$$\overline{R}^{(t)}(v_i) = \frac{1}{\deg(v_i)} \sum_{v_j \in \nu_K(v_i)} R^{(t)}(v_i v_j)\,. \tag{6.40}$$

This is effectively a Ricci curvature defined for an individual vertex, obtained through averaging over the curvatures between its K nearest-neighbors. Thus the distribution of vertices induces a distribution over $\overline{R}$. We denote such distributions by $p_f\left(\overline{R}; n\right)$, where f is the flavor label and n is the time step. We display such distributions in Fig. 6.9, where the top two panels display the overall unlabeled distribution after zero (left) and fifty (right) iterations of Ricci flow, while the bottom two panels depict the underlying distributions colored according to their flavor.

We see that the averaged-curvature distribution after fifty iterations of Ricci flow demonstrates a pronounced peak at the lower tail of its range, providing a clear value for a cut that separates the peak from the remainder of the distribution. In performing this cut and assessing the resultant accuracy, we achieve nearly the same level as with the UMAP cut: 91.13%.

What is particularly noteworthy about the distribution of averaged curvatures after fifty iterations of Ricci flow is the fact that the top jet distribution is strikingly localized while that of QCD jets is nearly-uniform. We can't help but posit that the latter is a reflection of the scale-free nature of QCD while the former has to do with the presence of the characteristic scale defined by m_{top}. Of course, any direct numerical relation is obfuscated by the drastic modification that fifty flow iterations impact on the distances between jets, so any substantive conclusions are certainly beyond the scope of this work. We believe concerted effort to connect these concepts would be very interesting indeed.

6.5 Discussion and conclusion

In this work, we have found the metric space defined by the EMD to be rather rich with information. By analyzing the space defined by samples of QCD- and top-initiated jets, we find jets of each class to fill their respective subspaces quite differently—in fact so differently that the labels of each jet can be determined purely through geometric information furnished by the EMD itself. The two methods pursued are (1) the density-based clustering through use of the DBSCAN algorithm and (2) the separation of subgraphs through use of the Ricci flow algorithm. Both of these methods are carried out completely unsupervised, achieve competitive accuracies, and each rely on only two initialization parameters. Furthermore, such parameters can be inferred from analyzing the data itself through physical reasoning. We compare some of the features to other leading top-taggers in Table 6.1.

As can be seen, our methods distinguish themselves not only through their unsupervised nature, but also with the $\mathcal{O}(1)$ number of parameters required for initialization. As the sophistication of ML techniques applied to problems of jet substructure evolve, we believe that features such as simplicity and explainability are to become ever more important. As physicists, we don't only

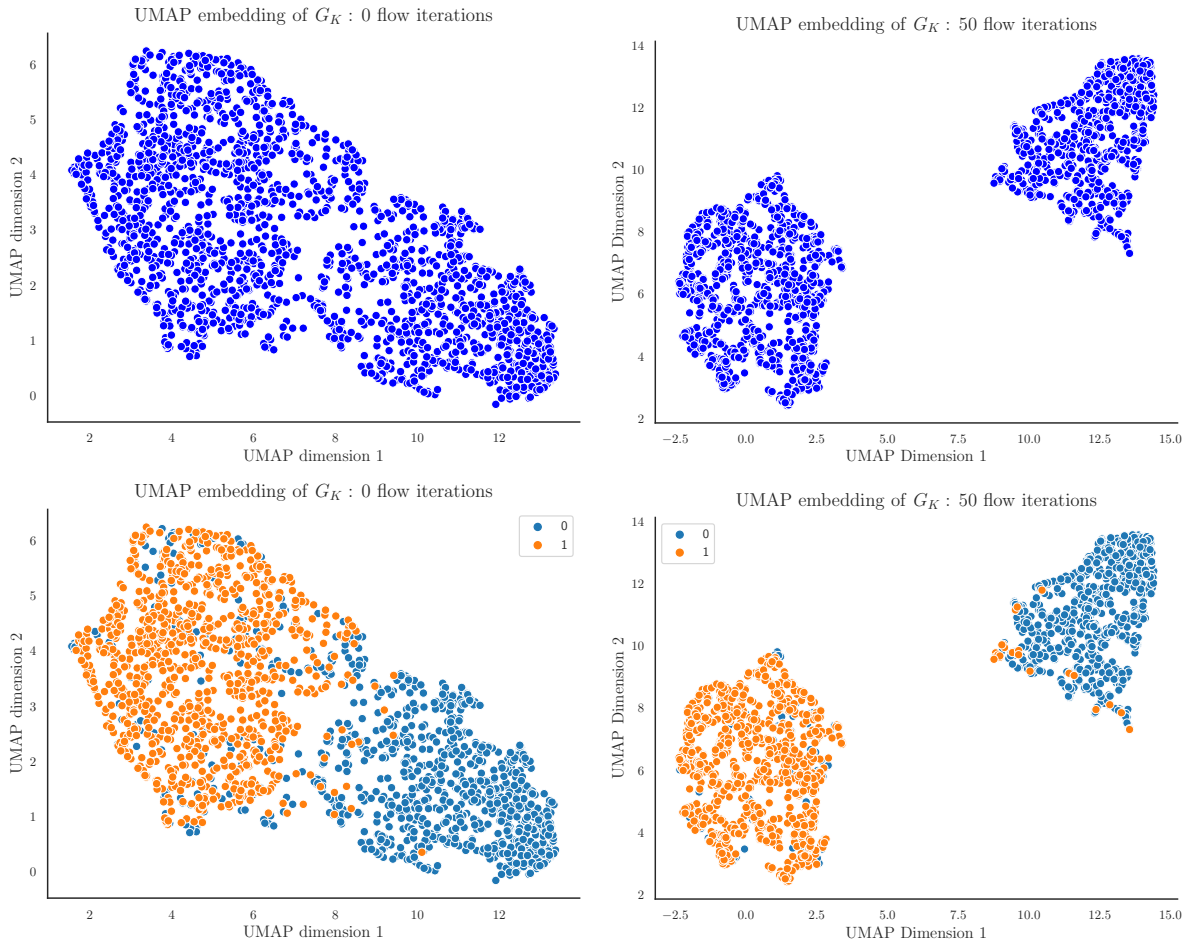

Figure 6.8: UMAP embeddings of the graph-path metric after zero (left) and fifty (right) iterations of Ricci flow. Results shown for unlabeled data above, with ground truth labels depicted below.

want tools that are effective in performing their tasks, but we also want tools whose effectiveness can be understood intuitively. Through our work, we see that such goals are certainly attainable and hope that the simple applications laid forth in this chapter can serve as a starting point for more refined studies. One immediate extension would be to the unsupervised tagging of jets initiated by other heavy resonances, such as $W/Z/H$, as the presence of large mass scales in each case should presumably lead to a fair degree of separation from QCD jets in EMD space. It would be also interesting to analyze the circumstances needed for the accurate clustering of jets in some low-dimensional space that the EMD manifolds are projected down to. In Ref. [PHO22], the authors developed a means of embedding manifolds whose distance is defined by the EMD into various

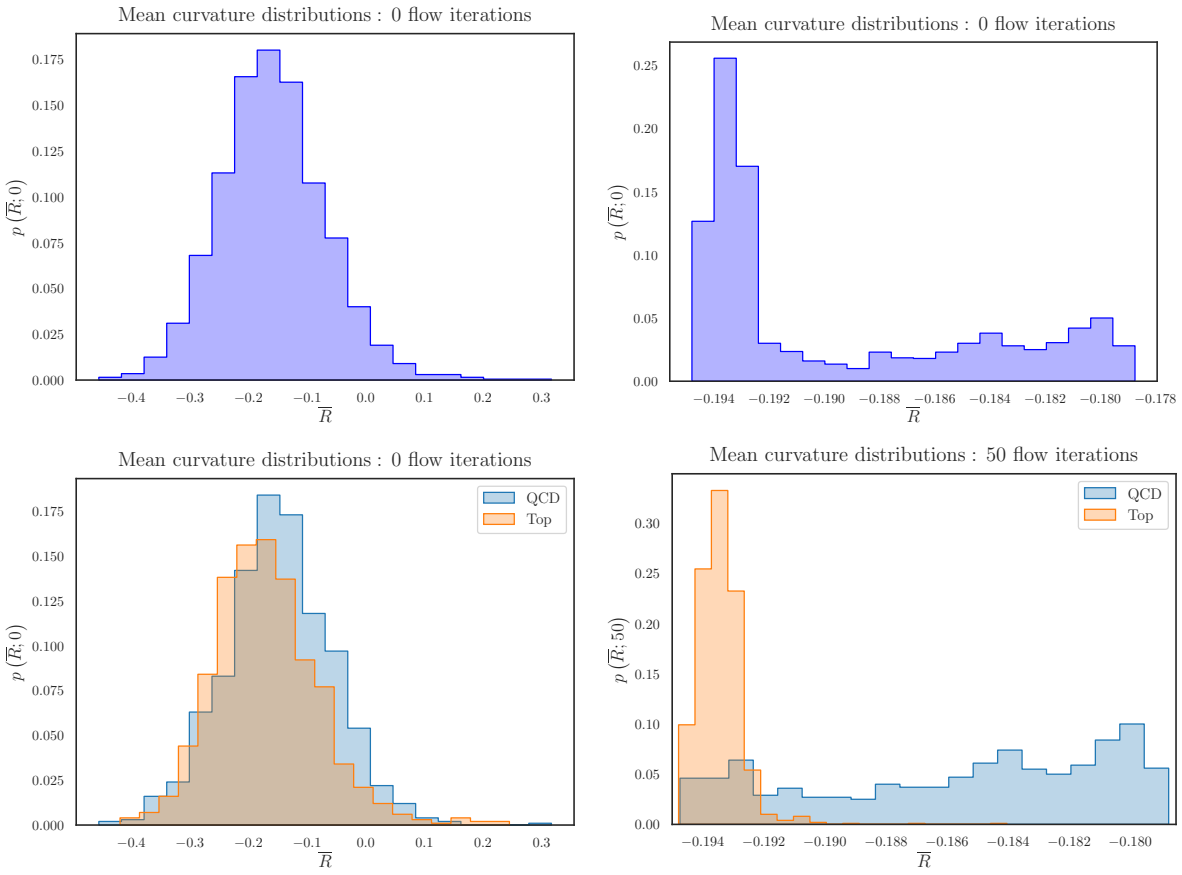

Figure 6.9: Averaged-curvature $\overline{R}$ distributions after zero (left) and fifty (right) iterations of Ricci flow. Results shown for unlabeled data above, with ground truth labels depicted below.

two-dimensional subspaces in a way that preserves as many of the features of the true manifold as possible—e.g. ascribing directions corresponding to jet mass and "pronginess." Carrying out a clustering analysis as a function of choice of two-dimensional embedding space would certainly be fascinating. We leave such investigations to future work.

Table 6.1: Comparison to a limited selection of top-taggers from the literature.

Architecture	**Accuracy**	**Parameters**	**Learning**
ResNeXt [XGD17]	0.9360	1.46e6	Supervised
ParticleNET [QG20]	0.9380	4.98e5	Supervised
PFN [KMT19b]	0.9320	8.20e4	Supervised
LGN [BAO20]	0.9290	4.50e4	Supervised
$\text{nPELICAN}_{\text{hidden=1}}$ [BHM23]	0.8951	11	Supervised
$\text{DBSCAN}_{\text{EMD}}$	0.9003	2	Unsupervised
$\text{Ricci-Flow}_{\text{Curvature}}$	0.9113	2	Unsupervised
$\text{Ricci-Flow}_{\text{UMAP}}$	0.9104	2	Unsupervised

CHAPTER 7

The simplicial substructure of jets

In this chapter, we construct a new data type for hadronic jets in which the traditional point-cloud representation is transformed into a simplicial complex consisting of vertices, or 0-simplexes. An angular resolution scale, r, is then drawn about each vertex, forming balls about hadrons. As r grows, the overlap of balls form 2- and 3-point connections, thereby appending 1- and 2-simplexes to the complex. We thus associate a jet with an angular-resolution-dependent characterization of its substructure—we dub this data type the simplicial substructure complex $K_{\rm sub}(r)$. This data type gives rise to two interesting representations. First, the subset of 0-and 1-simplexes lends itself naturally to a graph representation of a given jet's substructure and we provide examples of valuable graph-theoretic calculations such a representation affords. Second, the subset of 1- and 2-simplexes gives rise to what is known as a Face-Counting-Vector, in topological combinatorics parlance. We explore information-theoretic aspects of the components of this vector, various metric properties which follow, as well as how this vector can be used to define new jet-shape observables. The utility of these representations is demonstrated in the context of the discriminating of jets initiated by light quarks and gluons from those initiated by tops.

7.1 Introduction

The field of jet substructure concerns itself primarily with understanding the complexity of the radiation patterns developed by partons emerging from high-energy particle collisions [LMN17]. As such, it has proven itself to be a fertile ground for the application of various statistical and Machine-Learning-(ML)-based tools. For a comprehensive catalogue of such applications, see [FN21]. Early on, it was shown that jets and their substructure can be intuitively visualized as images [CKS15, OKM16], just as a calorimeter "views" them. Using this data type, one must perform a sequence of pre-processing steps to get the image in a standard format so that machines can learn the subtle differences among classes of jets, typically categorized according to the flavor of their initiating parton [KMS17, KPR17]. Such pre-processing amounts to the application isometries in the (η, ϕ) plane, physically corresponding to Lorentz boosts along the beam pipe. See [RW23] for a recent work that encodes the rotational symmetry exhibited by jets into neural network architectures.

Another important direction is the application of concepts from Optimal Transport [Vil09] to define a natural metric on the space of jets—see [KMT19c, KMT20] for the original ideas and [CCC20, CCC22] for extensions. Such a metric can be defined for jets in the discretized image format or the infinitesimal form known as a point-cloud. However, both representations still depend on the way in which jets are pre-processed, causing the "distance" between jets to inherit this dependence. Such an inheritance is ameliorated through use of an alternative data type, that is the spectral function [LT23], closely related to energy-energy correlators [BBE78, Tka97, JL11, LST13]. This data type makes use of inter-particle distances, thereby making its affiliated metric invariant under any pre-processing scheme.

Yet another interesting avenue is the application of concepts from Topological Data Analysis [CM21] to the study of jet substructure. This is carried out in [LLX20, TDK22], where jets are analyzed through topological notions of connectivity, such as persistent homology—see [LN22, HDP22, Beu23] for other applications of such notions to the physics of high-energy particle collisions. The basic structure underlying such analysis is that of a simplicial complex—that is a collection of points, lines, and triangles, or equivalently 0-, 1-, and 2-simplexes, where points

represent a particle with no connections, lines for two-particle connections, and triangles for three. Connections are activated based on the overlap of closed balls drawn about particles in the (η, ϕ) plane, and thus depend solely on inter-particle distances, making a simplical complex manifestly pre-processing independent. What is particularly novel about this construction is that a simplicial complex furnishes a one-parameter family of representations of a given jet, one for each value of angular resolution scale r—the radius of the ball drawn about each jet constituent.

The aim of this chapter is to unify the two aforementioned approaches to jet substructure—that is to represent a jet in a data type that (1) can be naturally endowed with a metric, (2) is completely independent of any choice of pre-processing, and (3) is a one-parameter family of representations, thus forming a thorough tomographic view of the various scales which characterize its substructure. We do this through defining a new fundamental data type, which we refer to as the simplical substructure complex of a jet, denoted $K_{\text{sub}}(r)$. This construction leads to two interesting limiting representations for jets, the first being a graph [Die17], which we label as $G_{\text{sub}}(r)$ and enables the computation of a host of graph-theoretical objects. The second representation is what is known as a Face-Counting-Vector (f-vector) [Zie95], denoted $\boldsymbol{f}(r)$, whose components are the number of 1- and 2-simplexes at a particular scale resolution scale r. The f-vector representation is particularly rich and the bulk of this chapter will be in the exploration of its features.

This chapter is organized as follows. Sec. 7.2 will be devoted to developing the data type and its limits, with Sec. 7.2.1 dealing with the construction of the simplicial substructure complex $K_{\text{sub}}(r)$, Sec. 7.2.2 the limiting case of the substructure graph $G_{\text{sub}}(r)$, and Sec. 7.2.3 the f-vector $\boldsymbol{f}(r)$. In Sec. 7.3 we examine geometric and information-theoretic consequences, as applied to the components of the f-vector, treated as random variables, in Sec. 7.3.1 and then with regard to QCD vs. top jets in Sec. 7.3.2. In Sec. 7.4 we demonstrate how the f-vector leads naturally to the definition of a new jet shape observable. We conclude in Sec. 7.5.

7.2 Jet substructure with simplicial complexes

In this section, we define the simplicial substructure complex for a jet, which we denote as $K_{\text{sub}}(r)$, where r is an angular scale that resolves the internal substructure of the jet. This is a fundamen-

tal object that naturally gives rise to two alternative jet representations, namely (1) a substructure graph, denoted $G_{\rm sub}(r)$ and (2) a substructure face-counting vector, $\boldsymbol{f}(r)$. These two representations may be viewed as particular limits of $K_{\rm sub}(r)$.

7.2.1 The simplicial complexes for a jet: basic definitions

The basic and natural data type for a jet is that of a point-cloud. A jet is a collection of the four-momentum vectors for the hadrons contained therein. The on-shell condition for final state particles then reduces the number of degrees of freedom from four down to three, which for hadron colliders like the LHC, are conveniently chosen to be the transverse momentum with respect to the beam pipe, p_T, together with the rapidity, η, and the azimuthal angle, ϕ. Then, normalizing the transverse momentum of each particle by that of the overall value for the jet, we obtain the fractions $z_i \equiv p_{Ti}/p_T^{\rm jet}$. The three degrees of freedom for each hadron are then (z_i, η_i, ϕ_i). With these three degrees of freedom, we construct the point-cloud representation for a jet

$$\mathcal{J}_{\rm pc}(\eta, \phi) = \sum_{i \in {\rm jet}} z_i \, \delta^{(2)}(\boldsymbol{\theta} - \boldsymbol{\theta}_i) \,, \tag{7.1}$$

where $\boldsymbol{\theta}_i \equiv (\eta_i, \phi_i)$. Hence, the point-cloud format represents the jet as a discrete probability distribution in rapidity-azimuth space with p_T-fractions as weights.

In this work, the base from which we start will be a simpler object, namely, the collection of $\boldsymbol{\theta}$-vectors. We will refer to this collection as $\mathcal{J}_\theta$, and let us also have $\mathcal{I} = \{1, 2, \ldots, N_{\rm jet}\}$ denote the indexing set so that

$$\mathcal{J}_\theta = \left\{ \boldsymbol{\theta}_i \in \mathbb{R}^2 \right\}_{i \in \mathcal{I}} \,, \tag{7.2}$$

where $N_{\rm jet}$ is the multiplicity of the jet. Note we are thus considering the jet as subset of the effective $\mathbb{R}^2$ formed by the calorimeter cells onto which the final state hadrons deposit themselves. Further note, that in this construction, we have completely dropped any z_i-dependence, treating particles purely in terms of their positions.

Next, for each particle in the jet, we can draw a closed ball about it, i.e. for the i^{th} particle, we construct the ball

$$B_i(r) \equiv \left\{ \boldsymbol{\theta} \in \mathbb{R}^2 \,|\, \|\boldsymbol{\theta} - \boldsymbol{\theta}_i\| \leq r \right\} , \tag{7.3}$$

where $\|\cdot\|$ denotes the Euclidean norm.

In what follows, since we are working with vectors in $\mathbb{R}^2$, we will consider simplexes up to order two, that is, 0-, 1-, and 2-simplexes. A 0-simplex will just be taken as a point $\boldsymbol{\theta}_i \in \mathbb{R}^2$ and denoted as $\langle \boldsymbol{\theta}_i \rangle$. A 1-simplex will be visualized as the line segment that connects two points $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ and denoted as $\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle$. Lastly, a 2-simplex can be visualized as the triangle formed by the connections between three points $\boldsymbol{\theta}_i$, $\boldsymbol{\theta}_j$, and $\boldsymbol{\theta}_k$, which is then denoted $\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \boldsymbol{\theta}_k \rangle$.

The generation of simplexes will come about through the varying of the radius r of the balls about each point in each jet uniformly. 1- and 2-simplexes are then spawned at the onset of nonempty intersections between pairs and triplets of balls, respectively. The number of 0-simplexes stays the same for all values of r, as it is just the set of vertices

$$S_0(r) = \{\langle \boldsymbol{\theta}_i \rangle\}_{i \in \mathcal{I}} \,. \tag{7.4}$$

As r grows, connections proliferate and one starts amassing collections of 1- and 2-simplexes. The set of the former is denoted

$$S_1(r) = \left\{ \langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle \,\middle|\, B_i(r) \cap B_j(r) \neq \emptyset \right\}_{i,j \in \mathcal{I}} \,, \tag{7.5}$$

where $\emptyset$ is the empty set. Thus, the condition for a connection between two points can be equivalently stated as

$$B_i(r) \cap B_j(r) \neq \emptyset \iff r \geq \frac{\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|}{2} \,. \tag{7.6}$$

The set of 2-simplexes is then

$$S_2(r) = \left\{ \langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \boldsymbol{\theta}_k \rangle \,\middle|\, B_i(r) \cap B_j(r) \cap B_k(r) \neq \emptyset \right\}_{i,j,k \in \mathcal{I}} \,. \tag{7.7}$$

For purposes of formality, let us define one last trivial set

$$S_{-1} = \emptyset \,. \tag{7.8}$$

Equipped with these sets, the simplicial substructure complex for a given jet is then defined as

$$K_{\text{sub}}(r) \equiv \bigcup_{\ell=0}^{3} S_{\ell-1}(r) \,. \tag{7.9}$$

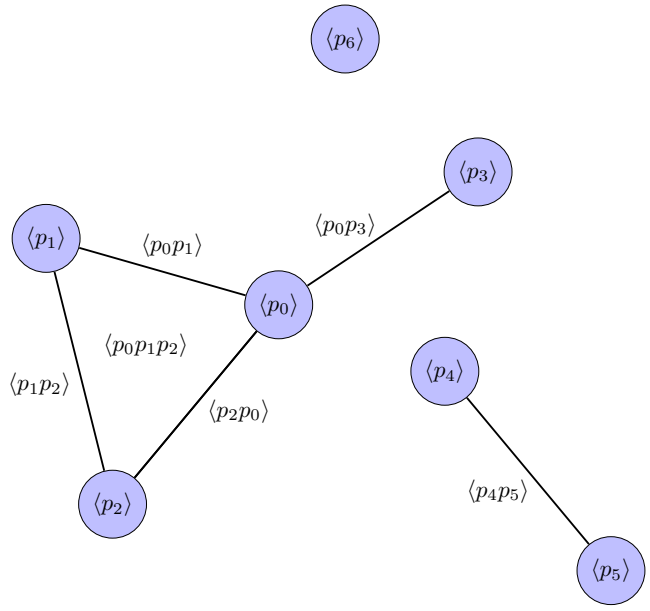

Figure 7.1: An example of a simplicial complex, $\mathcal{C}$. Here we have that $S_{-1} = \emptyset$, $S_0 = \{\langle p_0 \rangle, \ldots, \langle p_6 \rangle\}$, $S_1 = \{\langle p_0 p_1 \rangle, \langle p_1 p_2 \rangle, \langle p_2 p_0 \rangle, \langle p_0 p_3 \rangle, \langle p_4 p_5 \rangle\}$, $S_2 = \{\langle p_0 p_1 p_2 \rangle\}$, and therefore $\mathcal{C} = \bigcup_{\ell=0} S_{\ell-1}$.

At this point, an important physical comment is in order. Note that the simplicial substructure complex of Eq. (7.9) is purely a function of the radial coordinate r, whose magnitude is determined by the relative angular separations in the calorimeter cell. As such, $K_{\text{sub}}(r)$ is naturally invariant under isometries of the plane. ISO $(\mathbb{R}^2)$-invariance is to say invariance under translations, reflections, and rotations in (η, ϕ)-space. From a physical standpoint, such transformations correspond to Lorentz boosts along the beam pipe (canonically taken to define the z-axis) and from a data-formatting viewpoint, this is to say that $K_{\text{sub}}(r)$ is independent of pre-processing choice [CKS15, OKM16].

Before proceeding, we describe the particular data set used in this work, as well as the jet-trimming procedure we preform on the jets contained therein. The data set is the standard one used for the benchmarking of top-tagging architectures [But19]. The set consists of signal top jets amongst a mixed background of jets initiated by light quarks and gluons. These jets are simulated with Pythia8 [Bie22] for collisions at $\sqrt{s} = 14$ TeV. Detector effects are simulated with Delphes [FDD14] with the ATLAS card. Jets of radius $R = 0.8$ are then clustered using the anti-k_T

algorithm [CSS08] through FastJet [CSS12]. Each jet in the data set is the leading jet of the event from whence it came and falls in the transverse momentum range of $p_T^{\rm jet} \in [550, 650]$ GeV. Jets in this data set contain around fifty or so particles each—the vast majority of these being exceedingly low in transverse momentum and hence capturing soft physics. In the entirety of our proceeding analysis, we implement a basic trimming procedure [KTW10] in order to capture the hard/collinear physical scales which dominate the aspects of substructure in which QCD and top jets differ. Our procedure consists of simultaneously implementing a particle transverse momentum cut of $p_T^{\rm cut} \approx$ 3-5 GeV as well as a multiplicity cut of $N = 10$. Thus, we order the particles in descending order of their transverse momentum and keep only the ten highest-p_T particles, so long as they are above $p_T^{\rm cut}$. This guarantees that our jets are not contaminated by soft physics as well as gives our jets a fixed value of $N_{\rm jet} = 10$, which, as we will see, allows us to make sharp statements regarding various limits. Furthermore, we find this multiplicity cut to be quite ideal in revealing the critical angular resolution scale over which QCD and top jets differ the most, that is, what we refer to as the characteristic decay angle of tops

$$\theta_{\rm cd} \equiv \frac{m_{\rm top}}{p_T^{\rm jet}} \,, \tag{7.10}$$

where $m_{\rm top}$ is the top mass.[1] This is the unique angular resolution scale that can be constructed purely out of the available mass-dimension-1 parameters of the data set. We will see that the r-dependence of $K_{\rm sub}(r)$ is inherently what allows $\theta_{\rm cd}$ to reveal itself in such a wide variety of contexts as the critical scale of our data set.

7.2.2 The graph representation of a jet

We can think of the simplicial substructure complex, given by Eq. (7.9), as a principal representation of a jet, through which there exist limits which give rise to unique ways to resolve the substructure of a jet. In this section, we will consider the particular limit that restricts us to the set of 0- and 1-simplexes. Doing so will give rise to an r-dependent graph representation for each jet

[1] Note that this is technically the same as the "dead-cone" angle $\theta_{\rm dc}$, however that nomenclature should be reserved for the suppression of soft gluon radiation off a heavy quark. This is notoriously difficult to measure for top quarks [MST16] due to their short lifetime. However, it is identical to the characteristic opening angle of the decay products of boosted heavy particles.

[Die17], which we will refer to as $G_{\text{sub}}(r)$. We begin with some graph-theoretic preliminaries.

A weighted graph is a collection of vertices, edges, and edge weights $G = (V, E, W)$ where V is the set of vertices, E the set of edges, and W the set of edge weights. Using the language developed in the previous section, we can understand a weighted graph as a particular limit of a simplicial complex. In particular, we can identify each vertex with a 0-simplex and each edge with a 1-simplex, thus

$$V(r) \equiv S_0(r)\,, \tag{7.11}$$

$$E(r) \equiv S_1(r)\,, \tag{7.12}$$

where we make explicit the fact that both V and E inherit r-dependence from our simplicial construction. Next, let us consider a weight function, ω, that applies a weight to each edge according to

$$\begin{aligned} \omega\colon E(r) &\longrightarrow \mathbb{R}^{\geq 0}\,, \\ \langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle &\longmapsto \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|\ . \end{aligned} \tag{7.13}$$

The weight set W can thus be understood as the image of the edge set E under ω, and hence inherits r-dependence therefrom: $W(r) = \omega(E(r))$.

With these three sets so defined, we see that the simplicial substructure complex naturally gives rise to the graph substructure complex

$$K_{\text{sub}}(r) \longrightarrow G_{\text{sub}}(r)\,. \tag{7.14}$$

At this point we reiterate that as r grows, the development of 0- and 1-simplexes gives rise to non-trivial connection structures within each individual jet. A key distinction between the $G_{\text{sub}}(r)$ representation and the reduced $\mathcal{J}_\theta$ representation is that for each pair $\boldsymbol{\theta}_i, \boldsymbol{\theta}_j \in \mathcal{J}_\theta$, there is inherently a connection between them defined by their Euclidean distance $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|$, whereas for the corresponding pair $\langle \boldsymbol{\theta}_i \rangle, \langle \boldsymbol{\theta}_j \rangle \in G_{\text{sub}}(r)$, the same connection of $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|$ between them is only ascribed as a weight once the edge $\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle$ is formed, which is to say only once $r \geq \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|/2$. This leads to a non-trivial evolution of connection patterns in the graph representation. Such connections can be characterized by defining a path metric on the graph, where a path is taken to be a

sequence of 0-simplexes that are connected through 1-simplexes:

$$d_{\text{path}}\left(\langle\boldsymbol{\theta}_i\rangle, \langle\boldsymbol{\theta}_j\rangle\right) = \min\left\{\sum_{k=0}^{n-1} \omega\left(\left\langle\boldsymbol{\theta}_{A_k}\boldsymbol{\theta}_{A_{k+1}}\right\rangle\right) \;\middle|\; \left\langle\boldsymbol{\theta}_{A_k}\boldsymbol{\theta}_{A_{k+1}}\right\rangle \in S_1(r)\,,\, \langle\boldsymbol{\theta}_{A_0}\rangle = \langle\boldsymbol{\theta}_i\rangle\,,\, \langle\boldsymbol{\theta}_{A_n}\rangle = \langle\boldsymbol{\theta}_j\rangle\right\}. \tag{7.15}$$

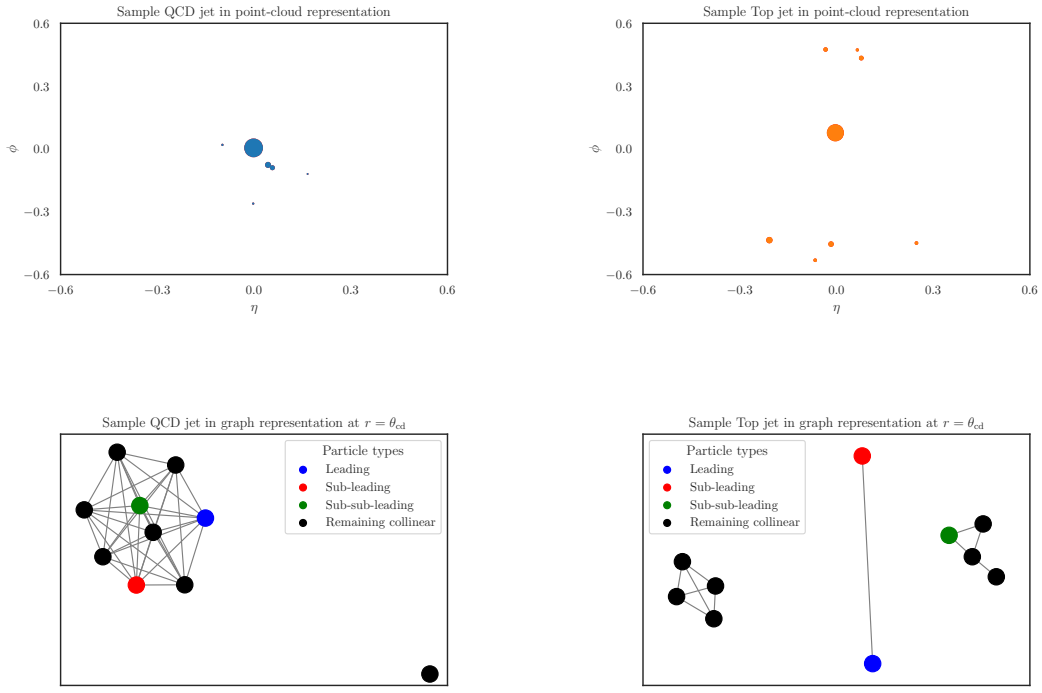

Figure 7.2: Examples of QCD (left) and top (right) jets in the point-cloud (top panels) and graph (bottom panels) representations. Note that in each case, our trimming procedure has already been enacted, leaving both jets with ten particles each. The graph representation is taken at the characteristic angular scale $r \simeq m_{\text{top}}/p_T^{\text{jet}}$. At this level of resolution, QCD and top graphs differ substantially in their both their numbers of connected components as well as the number of nodes contained therein.

With this metric in hand, we can ask questions regarding the global connectivity structure of a jet, and how this structure evolves with r. One particular quantity that is defined on the set of

vertices $V(r)$ is known as the "closeness centrality"

$$C_B(\langle\boldsymbol{\theta}\rangle) = \left(\frac{1}{N_{\text{jet}}}\sum_{\langle\boldsymbol{\theta}_i\rangle\sim\langle\boldsymbol{\theta}\rangle} d_{\text{path}}\left(\langle\boldsymbol{\theta}\rangle, \langle\boldsymbol{\theta}_i\rangle\right)\right)^{-1}, \tag{7.16}$$

which can be seen as an inverse average distance that the 0-simplex $\langle\boldsymbol{\theta}\rangle$ is from all other 0-simplexes that it is path-connected to, thus the smaller this average distance is, the closer $\langle\boldsymbol{\theta}\rangle$ is to all other points, and thus the more central of a location it takes within the overall graph.

In order to affiliate a single value of C_B with a jet, we consider the mean value taken over all vertices

$$\overline{C_B}(r) = \sum_{\langle\boldsymbol{\theta}_i\rangle\in G_{\text{sub}}(r)} C_B(\langle\boldsymbol{\theta}_i\rangle)\,. \tag{7.17}$$

Distributions for these mean closeness centralities for QCD and top jets for two different levels of angular resolution are depicted in Fig. 7.3. In this figure, we see that at the characteristic scale for the tops, QCD jets, by and large exhibit higher values than top jets, which is to say that the center of a QCD jet is predominantly close to its neighbors, whereas that of a top jet is, on average, much further from its neighbors. This is a graph-theoretic way of probing the following physical fact: the top jets have characteristic decay products which are boosted and spread out, on top of which QCD evolution takes place, adorning each prong with a haze of radiation. Thus, the center of the jet is inevitably far from all the haze spawned by each of the different decay products. Since QCD jets are predominantly single-pronged, the resulting haze surrounds this prong and the center itself. Thus, the mean closeness centrality evaluated at $r = \theta_{\text{cd}}$ is able to capture the connected/disconnected structure depicted in Fig. 7.2 for QCD/top graphs, respectively. This is confirmed by the distributions resulting from an angular resolution of $r = \theta_{\text{max}}$, where both QCD and top jets achieve full inter-connectivity among their vertices, and collapse to the same high value of $\overline{C_B}$.

To complement the global connectivity structure related to the elements of $S_0(r)$, we next look to the local structure that we may infer from elements of $S_1(r)$. To do so, we compute the Ricci curvature for a graph, known as the Forman curvature [SMJ16], which is defined on the set of

1-simplexes as

$$R_F\left(\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle\right) = 2 - \sum_{k \sim i, \ell \sim j} \left(\sqrt{\frac{\omega\left(\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle\right)}{\omega\left(\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_k \rangle\right)}} + \sqrt{\frac{\omega\left(\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle\right)}{\omega\left(\langle \boldsymbol{\theta}_j \boldsymbol{\theta}_\ell \rangle\right)}} \right) , \tag{7.18}$$

where the sum $k \sim i$ and $\ell \sim j$ are over immediate neighbors k of i and ℓ of j. This is to say paths that include only a single segment, which reduces $d_{\text{path}}\left(\langle \boldsymbol{\theta}_i \rangle, \langle \boldsymbol{\theta}_j \rangle\right) \to \omega(\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle)$, giving this particular quantity a notion of locality.

Just as we did with the closeness centrality, we ascribe to each jet its mean value of Forman curvature, taken over all edges in each $G_{\text{sub}}(r)$:

$$\overline{R_F} = \sum_{\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle \in G_{\text{sub}}(r)} R_F(\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_j \rangle) \, . \tag{7.19}$$

Fig. 7.4 displays distributions of the mean Forman curvatures for QCD and top jets at the same two angular resolution scales as in Fig. 7.3—namely $r = \theta_{\text{cd}}$ and $r = \theta_{\max}$. From this plot, we see that the mean Forman curvature behaves in a way that is inverse with respect to angular resolution scale to the behavior of the mean closeness centrality. We can understand this to arise from the sum appearing in Eq. (7.18), which has taken over only the immediate neighbors of vertices. Again, through the intuition garnered in Fig. 7.2, at $r = \theta_{\text{cd}}$ the disconnected structure of the top jets essentially collects three prongs over which to obtain mean curvatures from, each prong capturing a beam of QCD radiation, and therefore cause the top distribution to take on values near those of the QCD. Conversely, once we reach $r = \theta_{\max}$, all vertices gain connections, and therefore all vertices become immediate neighbors to all others. Thus, the sum over neighbors appearing in Eq. (7.18) becomes saturated. In going from θ_{cd} to $\theta_{\max}$, few neighbors are added to this summation in the case of QCD jets, thus resulting in only a minor additional spread to its distribution. This is to be contrasted with the case of top jets, whose vertices not only acquire far more neighbors over which to sum, but also the large edge weights which come with these very neighbors, characteristic of the far-extended angular substructure of top jets relative to QCD.

We thus see that the jet representation furnished by $G_{\text{sub}}(r)$ affords us the ability to compute graph-theoretic quantities, such as the closeness centrality and Forman curvature, which in turn, shed light on local/global features of the connectivity structure of jets probed at differing levels of angular resolution. We hope this graph representation opens the door to the exploration of many

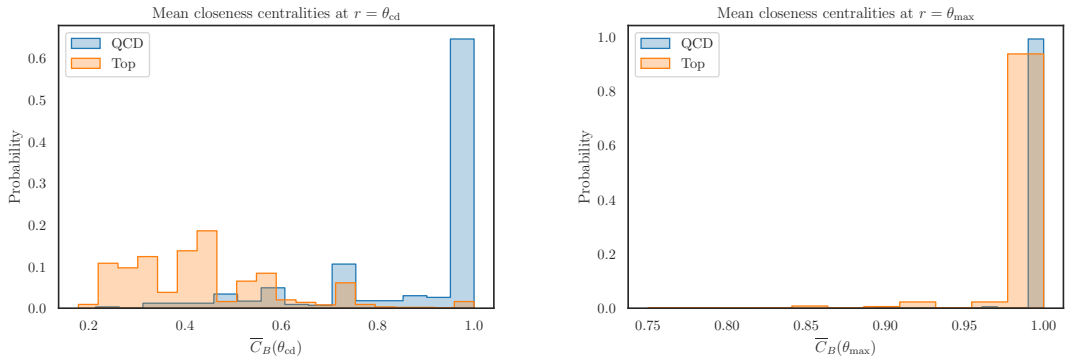

Figure 7.3: Closeness centrality distributions for QCD (blue) and top (orange) jet graphs evaluated at two angular resolution scales. The left plot is evaluated at the characteristic decay angle $r = \theta_{\mathrm{cd}}$ while the right is evaluated at the maximal angular resolution scale $r = \theta_{\mathrm{max}}$. We see that the behavior of closeness centrality distributions is largely dependent on the connectivity of the underlying graph, as at $r = \theta_{\mathrm{max}}$ both QCD and top jets are fully-connected and have their distributions collapse nearly onto one-another. This is to be contrasted with the case at $r = \theta_{\mathrm{cd}}$, where clear separability is achieved between the distributions.

other graph-theoretic concepts which can provide a helpful perspective from which to view jet substructure.

7.2.3 The f-vector representation of a jet

As alluded to previously, various limits of the simplicial substructure complex $K_{\mathrm{sub}}(r)$ highlight different aspects probed by the angular resolution scale r. The last section dealt with aspects of 0- and 1-simplexes, while the purpose of this section is to shed light on important features regarding the 1- and 2-simplexes. An obvious feature common to both $S_1(r)$ and $S_2(r)$ is that their cardinalities are monotonically-increasing functions of r, whereas that of $S_0(r)$ is constant and thus independent of r. To characterize such aspects, we define the cardinalities of each subset of simplexes forming $K_{\mathrm{sub}}(r)$:

$$f_\ell(r) \equiv \left| S_\ell(r) \right|, \tag{7.20}$$

where, by convention, we take $f_{-1}(r) = 1$ and we see trivially that $f_0(r) = N_{\mathrm{jet}}$.

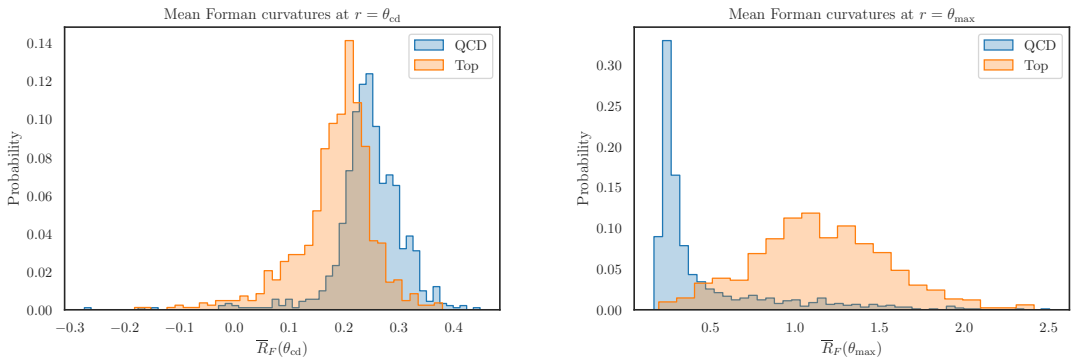

Figure 7.4: Forman curvature distributions for QCD (blue) and top (orange) jet graphs evaluated at two angular resolution scales. The left plot is evaluated at the characteristic decay angle $r = \theta_{\text{cd}}$ while the right is evaluated at the maximal angular resolution scale $r = \theta_{\text{max}}$. We see that the behavior of the Forman curvature distributions is somewhat opposite to those of the closeness centrality in that QCD and top distributions nearly coincide at scale $r = \theta_{\text{cd}}$ while they are highly separable at $r = \theta_{\text{max}}$. This is a manifestation of the Forman curvature local nature, which only differentiates top from QCD jets when the entirety of the extended structure of its decay products is resolved.

This definition gives rise to the construction of what is known as the face-counting vector [Zie95], or f-vector for short, for a given complex:

$$\boldsymbol{f}(r) = \big(f_{-1}(r), f_0(r), f_1(r), f_2(r)\big) \in \mathbb{N}^4 \,. \tag{7.21}$$

The components of the f-vector can equivalently be understood as arising from the following generating function, known as the f-polynomial[2]

$$f(t) = \sum_{\ell=0}^{d} f_{\ell-1}\, t^{\ell} \,. \tag{7.22}$$

Note that by studying jets of fixed multiplicity we automatically know the maximum cardinalities of each simplex set S_ℓ, which is to say we know the limits as $r \gg R$ where R is the overall jet radius. The limit is given simply by

$$\lim_{r \gg R} S_{\ell-1}(r) = \binom{N_{\text{jet}}}{\ell} \,. \tag{7.23}$$

This brings us to the key feature of this particular limit of $K_{\text{sub}}(r)$. By normalizing all the jets in a data-set to have the same fixed multiplicity—according to the procedure defined previously—we

[2] There exists a related polynomial, referred to as the h-polynomial

$$h(t) = \sum_{\ell=0}^{d} h_\ell\, t^{\ell} \,,$$

where the two are related via

$$\sum_{\ell=0}^{d} f_{\ell-1}\,(t-1)^{d-\ell} = \sum_{\ell=0}^{d} h_\ell\, t^{d-\ell} \,.$$

Upon expanding the above expression, one may identify h_3 with the reduced Euler Characteristic

$$\begin{aligned} h_3 &= -f_{-1} + f_0 - f_1 + f_2 \,, \\ &= -1 + V - E + F \,, \\ &= -1 + \chi \,, \\ &= \chi_{\text{red}} \,, \end{aligned}$$

where V is the number of vertices (f_0), E is the number of edges (f_1), and F is the number of faces (f_2) in the classical notation of polyhedra; χ being the usual Euler characteristic.

One may also extract the reduced Euler characteristic from the Euler-Poincaré formula for the f-vector

$$\chi_{\text{red}} = \sum_{\ell=0}^{d} (-1)^{\ell-1} f_{\ell-1} \,.$$

By the same token, the h-polynomial implies the existence of the h-vector of its coefficients.

see that the f-vector of Eq. (7.21) has two constant components and can thus be dimensionally-reduced to

$$\boldsymbol{f}(r) \to (f_1(r), f_2(r)) \in \mathbb{N}^2 \hookrightarrow \mathbb{R}^2 \,. \tag{7.24}$$

We will refer to this representation as the Simplicial Face-Counting Vector, or SFV for short. Note that by considering $\mathbb{N}^2$ to be embedded in $\mathbb{R}^2$, we can define the metric on the f-vectors to simply be that of the Euclidean metric induced by $\mathbb{R}^2$ under this embedding. As such, we can immediately visualize a sample of jets as living in the effective plane, which we denote as $\mathbb{R}^2_{\mathrm{S}}$. As r evolves, $\boldsymbol{f}(r)$ changes the locations of jets in $\mathbb{R}^2_{\mathrm{S}}$, and such motion can be visualized in Fig. 7.5. In describing this emergent space, we will use notation $\mathbb{R}^2_{\mathrm{S}} = \mathbb{F}_1 \times \mathbb{F}_2$ so that $(f_1(r), f_2(r)) \in \mathbb{F}_1 \times \mathbb{F}_2 \equiv \mathbb{R}^2_S \subset \mathbb{R}^2$, in order to distinguish the respective subspaces occupied by each component. We remark that $\mathbb{R}^2_S \subset \mathbb{R}^2$ in the strict sense, as

$$\binom{n(r)}{2} < \binom{n(r)}{3} \,, \tag{7.25}$$

where $n(r)$ is the effective number of 0-simplexes giving rise to 1- and 2-simplexes at the scale r. This inequality implies that $\mathbb{R}^2_S$ can be approximately intuited as the lower right triangular region contained in the rectangle spanned by

$$f_1 \in \left[0, \binom{N_{\mathrm{jet}}}{2}\right] \,, \quad f_2 \in \left[0, \binom{N_{\mathrm{jet}}}{3}\right] \,. \tag{7.26}$$

Displayed in Fig. 7.5 are kernel-density-estimate (KDE) plots depicting the distributions of QCD and top jets in this new space at three different levels of angular resolution—each level of resolution made with reference to θ_{cd}. The lowest level is $\theta_{\mathrm{cd}}/4$, where the total number of 1- and 2-simplexes is far below their maximum value thus both QCD and top jets occupy the lower left region of the space. We note that the QCD jets are smeared out over a larger range of values due to their predominantly single-pronged nature, while conversely, the higher degree of localization as manifested by the tops is due to their multi-pronged structure. The highest of the three resolution scales is chosen to be $2 \times \theta_{\mathrm{cd}}$, for which the vast majority of QCD jets begin to saturate their limits, while top jets are still smeared out over a larger expanse in their accumulation of simplexes. The intermediate scale is taken to be θ_{cd} where we see, as is to be expected by now, the greatest

separation between the two classes. At this characteristic angular scale of top jets, simplexes defined over their multiple prongs have largely saturated, and above this scale, simplexes defined across the various prongs begin to accumulate and the final approach to the limiting distribution can be made. This is to be contrasted with QCD jets, who by this point, have very-nearly saturated their limits. In Fig. 7.6, we replicate the projections onto the $\mathbb{F}_1$ and $\mathbb{F}_2$ spaces for a closer look at the separation between the distributions at the scale $\theta_{\rm cd}$.

We remark that a noteworthy feature of this construction is the very fact that $\mathbb{R}^2_S \subset \mathbb{R}^2$ and thus our embedding is readily visualized. Due to this, we can certainly classify QCD and top jets accurately by simply employing a histogram cut by eye on either of the projections in Fig. 7.6. Though alternatively, our Euclidean embedding allows for us to employ the standard unsupervised clustering algorithm, known as K-means [PVG11], in order to tag tops from the QCD background in $\mathbb{R}^2_S$ at the scale $r = \theta_{\rm cd}$. By simply recognizing the fact that we have a clear bimodal distribution at this scale, we can initialize K-Means to search for two clusters. Denoting our pair of clusters as $\mathcal{C} = \{C_1, C_2\}$, this amounts to the minimization task

$$\underset{\mathcal{C}}{\operatorname{argmin}} \sum_{a=1}^{2} \frac{1}{|C_a|} \sum_{\boldsymbol{f}_i, \boldsymbol{f}_j \in C_a} \left\| \boldsymbol{f}_i(\theta_{\rm cd}) - \boldsymbol{f}_j(\theta_{\rm cd}) \right\|^2 , \tag{7.27}$$

which, upon completion, clusters jets with 84% accuracy.[3]

7.3 Further investigations of the SFV

In this section, we delve further into properties of the SFV representation, but first we must develop some of the language required to do so. In what follows, we consider $f_1(r)$ and $f_2(r)$ as random variables whose stochasticity is inherited directly from that of our underlying data set. This is to say that the random fluctuations exhibited in the initial point-clouds of the QCD and top jets forming our data set are deterministically mapped to the components of the SFV through Eq. (7.20). This allows us to analyze properties regarding probability distributions in $f_1(r)$ and $f_2(r)$, where normalization comes simply from the total number of jets in each class and class

[3]This accuracy is determined by the F_1 score, a standard metric through which to assess the efficacy of unsupervised clustering algorithms.

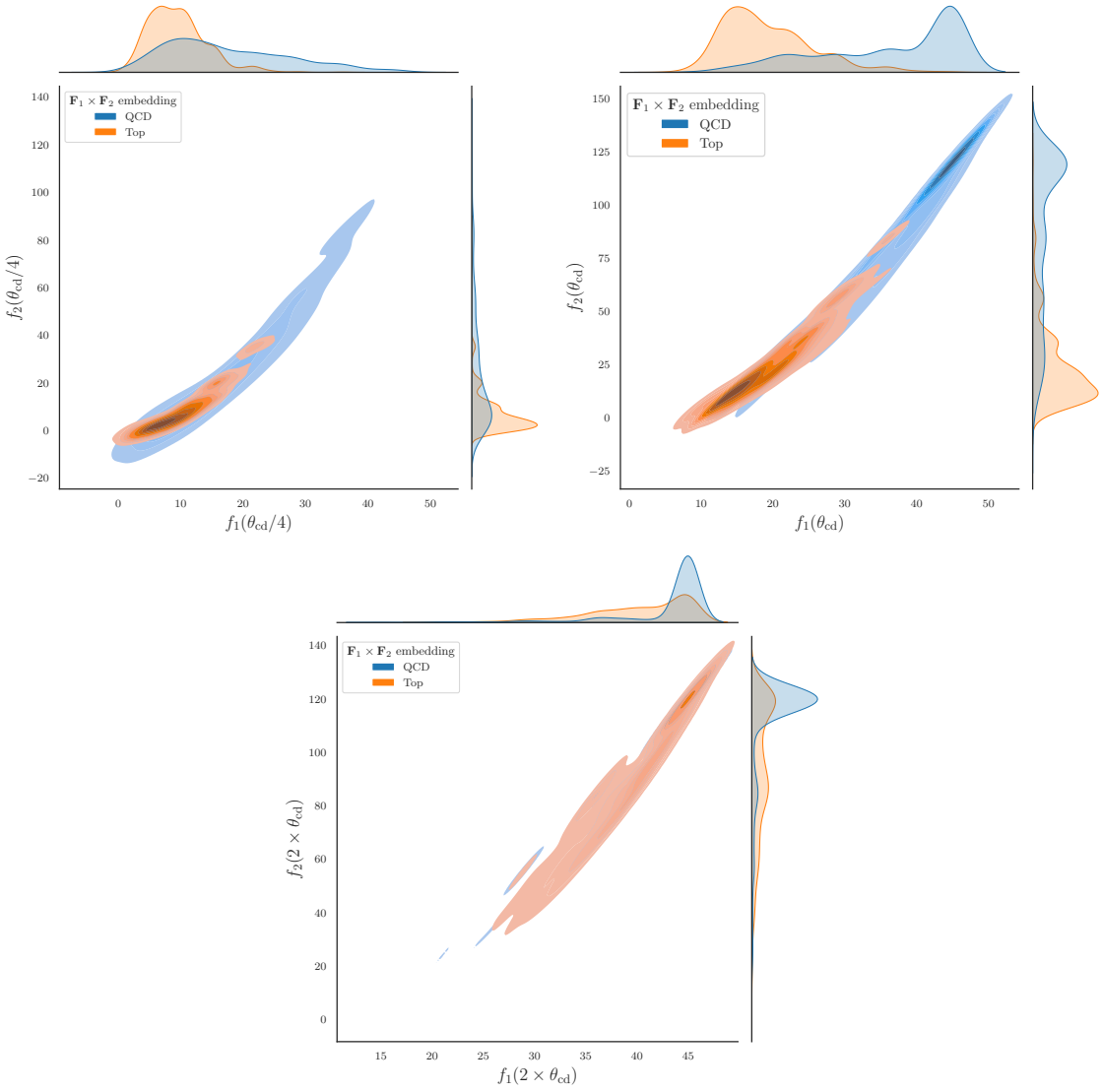

Figure 7.5: KDE plots for the embeddings of QCD (blue) and top (orange) jets in the $(f_1, f_2) \in \mathbb{F}_1 \times \mathbb{F}_2 \subset \mathbb{R}^2$ space at varying levels of angular resolution. The upper plots are generated at $r = \theta_{cd}/4$ (left) and $r = 2 \times \theta_{cd}$ while the lower plot corresponds to $r = \theta_{cd}$. We see the highest level of separability in the latter case.

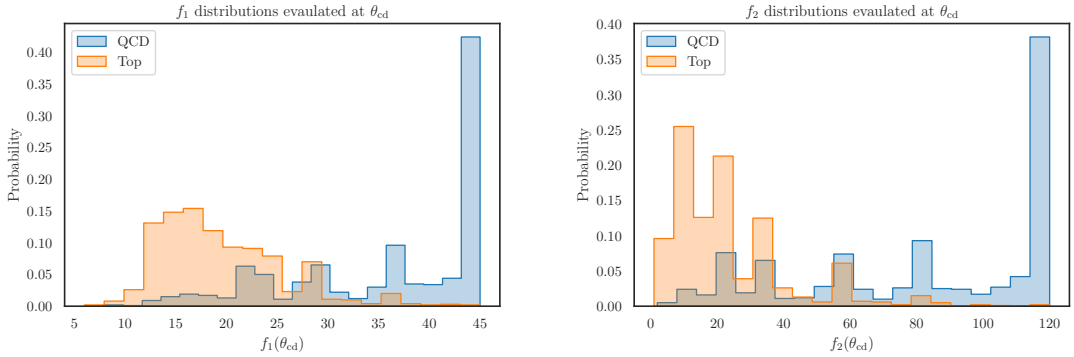

Figure 7.6: Distributions for f_1 and f_2 at the angular resolution scale $r = \theta_{\text{cd}}$. These may be viewed as projections onto the horizontal and vertical axes of the lower subfigure of Fig. 7.5.

membership is denoted by the subscript $A \in \{\text{QCD, top}\}$. Thus, for each jet flavor A and angular resolution r, there exists a one-parameter family of joint distributions in the the variables f_1 and f_2, $p_A\left(f_1, f_2; r\right)$.

With these considerations in mind, let us establish notation. We will follow that of [CT06]. We will refer to components of the SFV as random variables through F_1 and F_2 that are drawn from the distribution $p_A\left(f_1, f_2; r\right)$— this is denoted

$$F_1\,, F_2 \sim p_A\left(f_1, f_2; r\right)\,, \tag{7.28}$$

where

$$p_A\left(f_1, f_2; r\right) = \Pr\Big\{F_1 = f_1 \text{ and } F_2 = f_2\,, \text{at scale } r\,, \text{and for flavor } A\Big\}\,. \tag{7.29}$$

Next, we define what are known as the alphabets for F_1 and F_2, which are the discrete set of values these variables can take on. They are simply

$$\begin{aligned}\mathcal{F}_1 &= \left\{0, 1, \ldots, \binom{N_{\text{jet}}}{2}\right\}\,,\\ \mathcal{F}_2 &= \left\{0, 1, \ldots, \binom{N_{\text{jet}}}{3}\right\}\,.\end{aligned} \tag{7.30}$$

In light of the previous section, we see that these discrete sets are those that are embedded into the continuum spaces, $\mathcal{F}_{1,2} \hookrightarrow \mathbb{F}_{1,2}$ and visualized in Fig. 7.5.

Next, the information of a given pair (f_1, f_2) at scale r and for jet-flavor A is defined as

$$H_A(f_1, f_2; r) = \log \frac{1}{p_A\left(f_1, f_2; r\right)}\,, \tag{7.31}$$

where the intuition goes as follows. The larger the logarithm of the inverse probability to measure the pair (f_1, f_2), the smaller the probability weight affiliated with such a measurement, and thus the more informative such a measurement is about the underlying distribution from whence it is sampled.[4] In this work, we use the natural logarithm, or log base e and therefore measure information in what are known as "nats" (usual "bits" are defined for log base 2).

Finally, the KL divergence (or relative entropy) between two distributions P and Q of random variable X with alphabet $\mathcal{X}$ is a measure of information gain, or inefficiency, in modeling the "true" distribution P by the "model" Q. While the KL divergence is a generalization of a squared-distance, it is not a metric, as it is not symmetric and does not satisfy the triangle inequality [CT06]. It is computed as

$$D_{\mathrm{KL}}\left(P \,\|\, Q\right) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} \,. \tag{7.32}$$

These basic ingredients will be utilized in what follows. First, in Sec. 7.3.1 we explore some basic information-theoretic and geometric considerations regarding the random variables F_1 and F_2 themselves and considering the flavor-dependence of such considerations separately. Then in Sec. 7.3.2, we will consider the information overlap and distances between distributions as indexed by their jet flavor A resulting from their underlying values of (f_1, f_2). We emphasize that the r-dependence inherited by the SFV from the simplicial substructure complex $K_{\mathrm{sub}}(r)$ affords us one-parameter families of distributions $p_A\left(f_1, f_2; r\right)$, which in turn allow us to study all the following information-theoretic and geometric features as functions of the angular resolution scale r.

7.3.1 Information and geometry of the random variables F_1 and F_2

Upon examining Fig. 7.5, we qualitatively see correlation between the f_1 and f_2 coordinates. Now, it is intuitively obvious that f_1 and f_2 should be correlated, and a natural information-theoretic object to quantitatively investigate such a correlation is the mutual information between the two

[4] This is why this quantity is often also referred to as the "surprise" of a measurement.

random variables F_1 and F_2. This is defined as

$$I_A\left(F_1, F_2; r\right) = \sum_{f_1 \in \mathcal{F}_1\,,\, f_2 \in \mathcal{F}_2} p_A\left(f_1, f_2; r\right) \log \frac{p_A\left(f_1, f_2; r\right)}{p_A\left(f_1; r\right)\, p_A\left(f_2; r\right)}\,, \tag{7.33}$$

or equivalently

$$I_A\left(F_1, F_2; r\right) = D_{\mathrm{KL}}\left(p_A\left(f_1, f_2\right) \,\middle\|\, p_A\left(f_1\right)\, p_A\left(f_2\right)\right)\,, \tag{7.34}$$

where

$$p_A\left(f_i; r\right) = \sum_{f_j \in \mathcal{F}_j} p_A\left(f_i, f_j; r\right)\,, \text{ for } i \in \{1, 2\}\,,\, j \neq i\,, \tag{7.35}$$

are the marginalized distributions. Eq. (7.34) thus tells us that the mutual information quantifies the extent to which the joint distribution in (f_1, f_2) deviates the product of its marginals. Such a product only well-approximates the joint distribution in the limit that F_1 and F_2 become independent.

A quantity closely-related to the mutual information is what's known as the variation of information—it is defined by

$$\mathrm{VI}_A\left(F_1, F_2; r\right) = S_A\left(F_1, F_2; r\right) - I_A\left(F_1, F_2; r\right)\,, \tag{7.36}$$

where $I\left(F_1, F_2; r\right)$ is given by Eq. (7.34) and $S_A\left(F_1, F_2; r\right)$ is the joint Shannon entropy

$$S_A\left(F_1, F_2; r\right) = -\sum_{f_1 \in \mathcal{F}_1\,,\, f_2 \in \mathcal{F}_2} p_A\left(f_1, f_2\right) \log p_A\left(f_1, f_2\right)\,, \tag{7.37}$$

We remark that the variation of information, $\mathrm{VI}_A\left(F_1, F_2; r\right)$ of Eq. (7.36), satisfies all the requirements for a distance metric. As such, it can be interpreted as the distance between the random variables F_1 and F_2, or in other words, the separation in the information contained in F_1 and F_2, at the scale r

With these objects defined, we may see how they behave as functions of the angular resolution for QCD and top jets separately, beginning with the mutual information displayed in Fig. 7.7. Based on the discussion surrounding Eq. (7.34), we may identify the regions in which $I_A\left(F_1, F_2; r\right)$ reaches a global/local maximum as the regions in which the F_1 and F_2 variables have the highest correlation. These are the very regions in which both 1- and 2-simplexes proliferate concurrently,

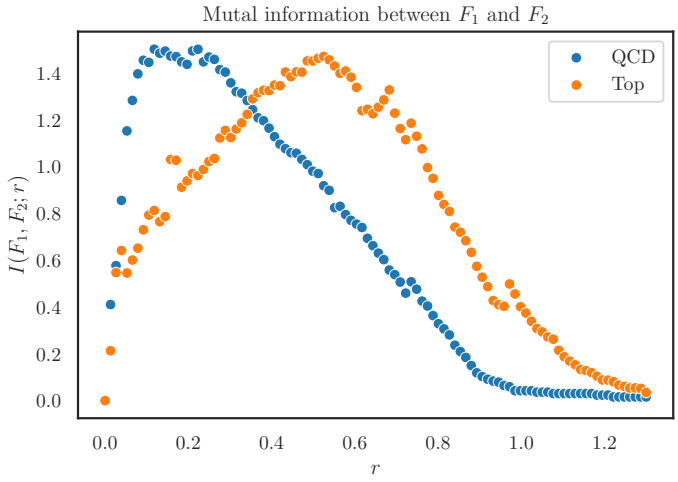

Figure 7.7: Mutual information for QCD (blue) and top (orange) jets. The mutual information captures in the information overlap between the random variables F_1 and F_2.

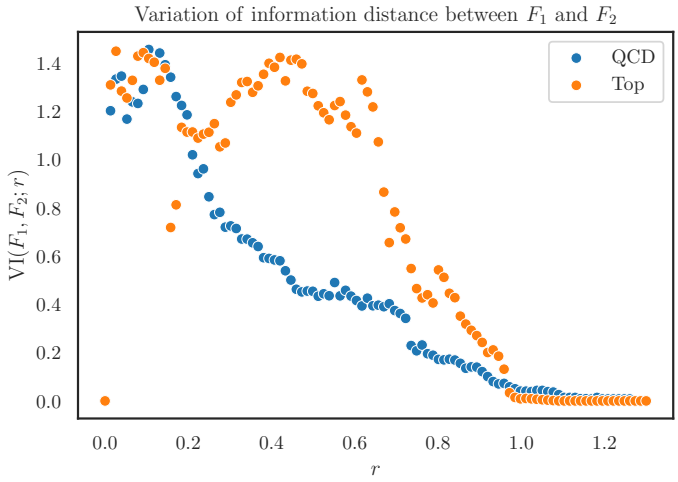

Figure 7.8: Variation of information for QCD (blue) and top (orange) jets. The variation of information is a measure of the distance between the random variables F_1 and F_2.

and such regions are those where dense clusters of particles are to be found, as each $B_i(r)$ has nonzero intersection with neighboring balls. Hence these are the regions in which the angular scale r resolves dense clusters.

Now, looking to Fig. 7.7, we may understand how this behavior manifests itself for both QCD and top jets. First, in the QCD case, we identify a maximum plateau for $r \sim [0.1, 0.2]$ after which $I_{\mathrm{QCD}}\left(F_1, F_2; r\right)$ falls steadily—thus identifying a dense core of collinear radiation about a single hard prong. This type of behavior is only experienced in $I_{\mathrm{top}}\left(F_1, F_2; r\right)$ once $r \gtrsim \theta_{\mathrm{cd}}$. This is a byproduct of the extended structure of top-initiated jets, due to the top quark's characteristic hard decay pattern of $t \to q\bar{q}'b$, whose extent is only fully-resolved for these high values of r. Thus, we see that in both cases the mutual information provides a gauge for the effective angular width of a jet's substructure, for it tracks the level of correlation between the proliferation of 1- and 2-simplexes. Such proliferation proceeds up the bulk of the particles are resolved, and then slowly decays down to zero once the values of f_1 and f_2 approach their limits. Alternatively, we may interpret the r-dependent mutual information to provide a proxy for the density with which particles fill their respective jets. The lower the value of r for which $I_A\left(F_1, F_2; r\right)$ attains its max, the more highly-collimated and dense the jets are, and vice versa.

The variation of information, $\mathrm{VI}_A\left(F_1, F_2; r\right)$, is shown in Fig. 7.8. Due to its close relation to $I_A\left(F_1, F_2; r\right)$, we may reasonably expect qualitatively similar behavior to exhibit itself, however, we see that this metric reveals some interesting sharper features. First, let us note the features this metric has in common with the mutual information, which are the approximate locations of the maximum for QCD jets near $r \sim 0.1$ and the max plateau region for top at $r \gtrsim \theta_{\mathrm{cd}}$.

The feature that is most striking is the emergence of two maxima of the same size for the case of top jets. Considering Figs. 7.7 and 7.8 together and inspecting Eq. (7.36), we see that this additional maximum originates in a local maximum in the joint entropy $S_{\mathrm{top}}\left(F_1, F_2; r\right)$, which must attain a lower value than its global max—we see that global maximum must occur at the same location as $I_{\mathrm{top}}\left(F_1, F_2; r\right)$ with value about twice that of the mutual information. The presence of this local max is due to the proliferation of simplexes in the smaller sub-clusters of particles that appear in top jets. We know these must form in subsets of the total number of particles contained in the jet, because the totality of the top jet's constituents can only be resolved at the higher values

of r.

Now, what makes the variation of information a particularly interesting metric is the physical interpretation of the peaks near $r \sim 0.1$ for both QCD and top jets. We can interpret these peaks to define the resolution scale for collinear radiation about hard colored prongs due to parton showering—the single prong for QCD jets and the triplet in the case of tops. Remarkably, we have

$$S_{\text{QCD}}\left(F_1, F_2; 0.1\right) \approx S_{\text{top}}\left(F_1, F_2; 0.1\right) \ , \tag{7.38}$$

despite the two jet classes resolving widely differing numbers of particles in their respective clusters—$\mathcal{O}(10)$ for QCD and $\mathcal{O}(1)$ for top. This implies that the distance between F_1 and F_2 at angular scales that resolving subjets. This is then further corroborated by

$$S_{\text{top}}\left(F_1, F_2; 0.1\right) \approx S_{\text{top}}\left(F_1, F_2; 0.5\right) \ , \tag{7.39}$$

where for $r \sim 0.5$, the balls surrounding particles contained in each of the three prongs of the top jets have large enough radii to form simplexes with those of neighboring prongs, and we enter the regime of maximal joint entropy. This regime can then be interpreted as the regime where r is too large to resolve the details of the particles forming each prong, but can now recognize the emergence of a new set of subjets to recluster and subsume into a larger one. Thus, we see the maxima of $\text{VI}_A\left(F_1, F_2; r\right)$ as indicators of the telescoping/fractal structure of jets. While QCD jets are scale-free and thus contain only one peak denoting continual self-similarity as r grows, the presence of the top mass introduces an additional scale, thereby introducing an additional peak in $\text{VI}_A\left(F_1, F_2; r\right)$, and hence signifies an additional layer of substructure through which to cluster subjets contained within top jets.

The fundamental differences between QCD and top jets are thus revealed quite dramatically by the SFV data type. In the following section, we demonstrate that the sensitivity of the SFV to absence/presence of extended structures may be leveraged to reveal distinguishing features of QCD and top jets.

7.3.2 Information and geometry of QCD and top jet distributions

In this section, we investigate how the SFV data type can be used to explicitly differentiate QCD and top jets. Such differentiation is the focus of many ML-based studies in the HEP literature, as top-tagging is a task of primary interest at colliders such as the LHC. The study of flavor relations between distributions can be cast into the physical language of top-tagging, as such a pursuit consists of one searching a background high-probability haze of QCD jets for anomalous low-probability top jets. Recalling our previous discussion, the KL divergence

$$D_{\mathrm{KL}}\left(p_{\mathrm{QCD}} \,\|\, p_{\mathrm{top}}; r\right) = \sum_{f_1 \in \mathcal{F}_1\,,\, f_2 \in \mathcal{F}_2} p_{\mathrm{QCD}}\left(f_1, f_2; r\right) \log \frac{p_{\mathrm{QCD}}\left(f_1, f_2; r\right)}{p_{\mathrm{top}}\left(f_1, f_2; r\right)}\,, \tag{7.40}$$

measures the inefficiency in modeling the distribution of background QCD jets by the distribution of the rare top jets. The resulting curve is displayed in Fig. (7.9), where the most interesting feature is the location of its peak. We note that our data set contains jets of $p_T^{\mathrm{jet}} \in [550, 650]$ GeV, and thus we see that

$$\underset{r}{\mathrm{argmax}}\, D_{\mathrm{KL}}\left(p_{\mathrm{QCD}} \,\|\, p_{\mathrm{top}}; r\right) \approx \theta_{\mathrm{cd}}\,. \tag{7.41}$$

Therefore, this quantity readily identifies the angular scale unique to tops.

While the KL divergence is a useful measure of differing information content for the two jet flavors, such analysis may be complemented by the study of a metric defined on the distributions themselves. One such metric is the Hellinger distance. For the case of QCD and top SFV distributions, the square of this distance is computed as

$$H^2(p_{\mathrm{QCD}}, p_{\mathrm{top}}; r) = \frac{1}{2} \sum_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} \left(\sqrt{p_{\mathrm{QCD}}(f_1, f_2; r)} - \sqrt{p_{\mathrm{top}}(f_1, f_2; r)} \right)^2 . \tag{7.42}$$

$H(p_{\mathrm{QCD}}, p_{\mathrm{top}}; r)$ is displayed in Fig. 7.10. The unique feature of the Hellinger distance is its effective identification of the two aforementioned angular resolution scales that appear in top jets. Note that Hellinger distance displays a single global maximum plateau defined sharply by the region $r \in [0.1, \theta_{\mathrm{cd}}]$. This highlights something that is seen neither in the analysis of Sec. 7.3.1, nor in that of the KL divergence of the present section. The fact that the angular scale $r \approx 0.1$ marks a boundary for the maximal distance between distributions means that at the level of collinear

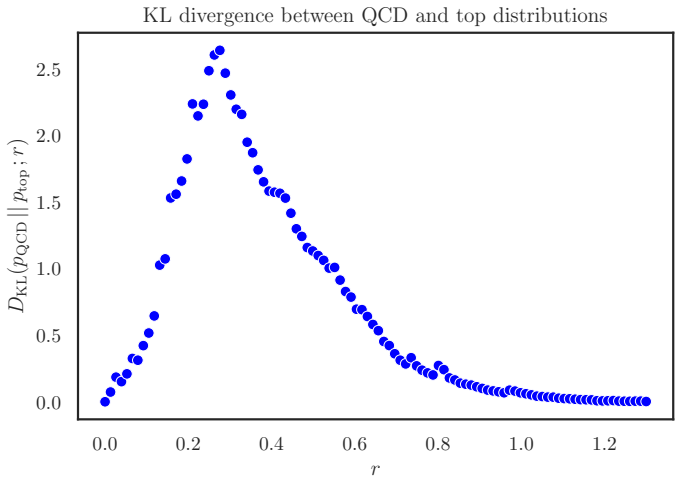

Figure 7.9: The KL divergence between QCD and top f_1-f_2-distributions. The KL divergence captures in the information-theoretic surprise one gets in modeling the top distribution by that of the QCD.

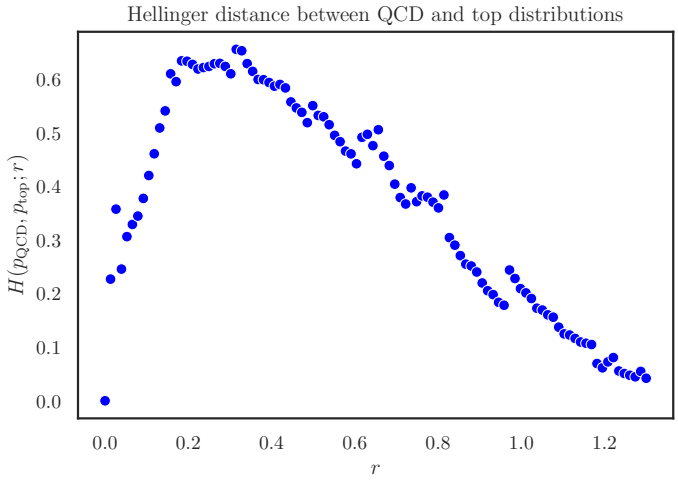

Figure 7.10: The Hellinger distance between QCD and top f_1-f_2-distributions. The Hellinger distance defines a metric on between these distributions.

radiation, the QCD and top distributions differ sharply in their values of $f_1(0.1)$ and $f_2(0.1)$, as at this scale, QCD jets have nearly the entirety of their particles resolved, whereas top jets have only the particles immediately surrounding each core resolved. Thus, QCD jets exist in their maximal delta function regime, whereas top jets still require a wider angular scale to capture the entirety of their particle content. This latter point is then the reason that the upper boundary of the Hellinger maximum lies at θ_{cd}—for once this scale is reached, neighboring cores are able to resolve each other's presence, and the top distribution thus begins its limiting to the delta function regime. Thus, the Hellinger distance is the best means of understanding the separability of QCD and top jets in $\mathbb{R}^2_{\text{S}}$ shown in Fig. 7.5.

7.4 Simplicial jet shape observable

In this final section, we define an experimental observable that follows naturally from the SFV data type. This observable is closely related to what are known as the integrated and differential jet shapes and is defined nearly identically.

The integrated jet shape is a classic jet substructure observable and can be understood as follows. An axis within the jet is chosen about which an angular scale r is chosen. The energies of the particles contained within the solid disk spanned by r are summed up and normalized by the overall energy of the jet. It may thus be interpreted as a cumulative distribution function in the angle r obtained through the first moment of the particle-energy-fraction, z, distribution. Thus letting z denote the fraction of energy carried within the radius r, so that z has implicit dependence on r, we have

$$\Psi(r) = \int_0^1 dz\, z(r)\, p(z(r))\,, \tag{7.43}$$

where $p(z(r))$ is the distribution of energy fractions z evaluated at the resolution r of a sample of jets. The differential jet shape is then simply the r-derivative of Eq. (7.43), giving it the interpretation of a probability density function with respect to r:

$$\psi(r) = \frac{d}{dr}\Psi(r)\,, \tag{7.44}$$

see [Sey97, Sey98, VWZ08b, CV14a, CV16b, KRW17, NPW19a, CRW19a] for various theoretical calculations of this observable.

We see that this definition lends itself quite naturally to an analogous jet shape observable with respect to the SFV. Here $\boldsymbol{f}(r)$ plays the role of $z(r)$ which is the accumulated number of simplexes contained within the angular resolution variable r. A key difference, though, is that $\boldsymbol{f}(r)$ makes no reference to a particular axis choice, as the angular scale r is that which defines balls about each particle contained within the jet. Thus, the pre-processing independence of the SFV leads to an independence in the choice of axis for the simplicial shape.[5] Let $\|\boldsymbol{f}(r)\| = \sqrt{f_1(r)^2 + f_2(r)^2} \equiv \mathcal{N}(f_1, f_2; r)$, and then, in direct analogy with Eq. (7.43), we define the integrated simplicial shape to be

$$\Sigma(r) = \sum_{f_1 \in \mathcal{F}_1\,,\, f_2 \in \mathcal{F}_2} \mathcal{N}(f_1, f_2; r)\, p\,(f_1, f_2; r)\ , \tag{7.45}$$

so that the differential version is directly analogous to Eq. (7.44), i.e.

$$\sigma(r) = \frac{d}{dr}\Sigma(r)\,. \tag{7.46}$$

Note that the cumulative distribution (probability density) function interpretations of the integrated (differential) shape is not being used here, as we are taking $\mathcal{N}(f_1, f_2; r)$ to bot be normalized by its maximum value. Our choice is simply one of many choices for simplicial shape observables.

Plots for $\Sigma(r)$ and $\sigma(r)$ for both QCD and top jets are displayed in Figs. 7.11 and 7.12, respectively. They display behavior that is congruent with the discussions of Sec. 7.3. At both the integrated and differential levels, we see the rapid accumulation of simplexes at the resolution scale $r \sim 0.1$ for QCD jets, while for top jets, this takes place initially at $r \sim 0.1$ where the QCD radiation surrounding each prong is resolved, and then again once $r \gtrsim \theta_{\text{cd}}$ and where the prongs can be resolved into a single large-scale cluster. Thus both $\Sigma(r)$ and $\sigma(r)$ are shape observables which are particularly well-suited for exhibiting the extended nature of decay patterns within the substructure of a jet. We remark here that the shapes displayed are simply for the particular choice of $\mathcal{N}(f_1, f_2; r)$ given above, and that there certainly exist many more functions of SFV components

[5] From a theoretical standpoint, computation of the traditional jet shape has a remarkably strong dependence on the choice of axis through remarkably different QCD factorization theorems that each axis requires [KRW17].

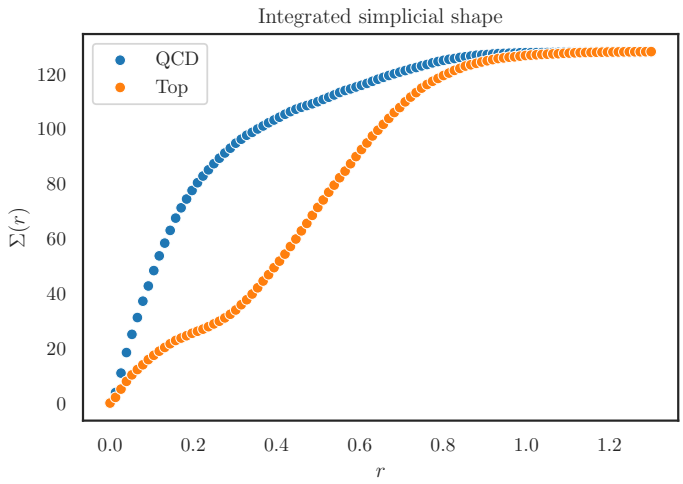

Figure 7.11: Integrated simplicial shapes for QCD (blue) and top (orange) jets.

one could choose—and in fact it would be very interesting so see if different choices are better suited for identifying other important physical features of a jet's substructure. On such a note, it would also furthermore be interesting to see the corresponding profiles for jets whose progeny comes from other heavy resonances, such as $W^{\pm}$, Z, or H bosons, for the identification of their respective extended structures could serve in the performance of precision measurements of the masses of such particles.

7.5 Conclusion

In this chapter, we have proposed a fundamentally new data type for the representation of jets, which we refer to as the simplicial substructure complex, $K_{\text{sub}}(r)$. This data type is the union of sets of 0-, 1-, and 2-simplexes, whose elements are dependent on the internal angular resolution variable r of a given jet. This data type gives rise to two natural representations, namely the graph representation $G_{\text{sub}}(r)$ and the SFV representation $\boldsymbol{f}(r)$. Each representation allows for the computation of novel features. The graph representation allows for the insights of graph theory to be applied to jet substructure, which is particularly useful in the study of the prong structure of jets. The SFV representation readily lends itself to visualization, analytic computation of distances between vectors, as well as a host of information-theoretic and geometric considerations at the

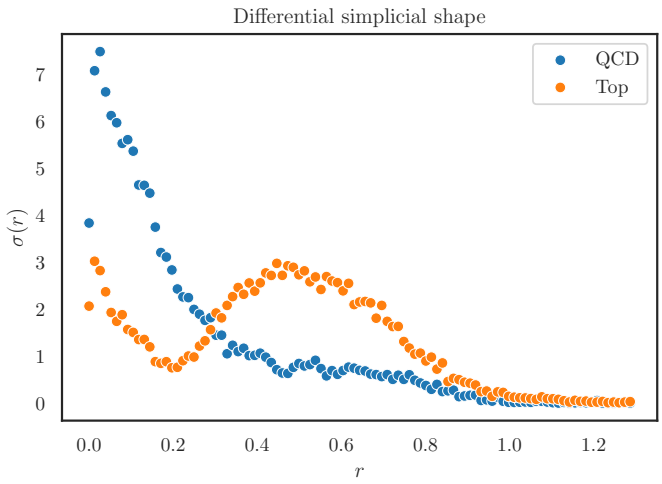

Figure 7.12: Differential simplicial shapes for QCD (blue) and top (orange) jets.

level of random variables and distribution flavors.

Thus, we see that the SFV distinguishes itself in many previously-mentioned regards, but we highlight one in particular—that is its scale-dependence. The analysis of such dependence is of central interest to both experimental and theoretical investigations of jet substructure and is physically an extremely natural dial through which to probe the internal workings of jets. This scale-dependence is also of fundamental importance in the field of topological data analysis, which certainly makes TDA seem to be a natural lens through which to view jet substructure. We hope that this work can serve to open the door to a greater union between these disciplines as well as many more interesting studies along this line.

CHAPTER 8

Conclusion

In this PhD thesis, we have explored one of the most interesting many-body emergent phenomena in nature, that is, the hadronic jets produced by the high-energy collisions of hadrons. In this pursuit, we have focused our attention on jets whose progenitors are heavy quarks, such as charm, bottom, and top. The distinguishing features of such quarks are their large masses, and it is a remarkable fact that these masses imprint themselves in subtle ways on the radiation pattern formed in their respective jets.

The first section of this thesis concerned itself with the measurement of back-to-back heavy-quark dijet pairs, and how such measurements can be utilized to infer important nuclear physics concerning the initial and final states of collisions. The first such study sought to probe the flavor-dependence of parton energy-loss in the QGP created in the collision of heavy ions. In doing so, we discovered that the invariant mass spectrum of b-dijets is an exceptionally sensitive probe of such energy-loss mechanisms, due to its overwhelming suppression in the heavy-ion context relative to the vacuum case. The second study demonstrated how one may use c- and b-dijets to directly probe the gluonic contribution of the Sivers asymmetry of the initial state proton in deep-inelastic scattering experiments. The former is of utmost importance for the sPEHNIX experiment while the latter pertains to the goals of the EIC.

Next, we peered into the internal substructure of jets initiated by charm and bottom quarks in order to reveal the precise imprinting of their masses on their radiation profiles, as alluded to previously. By considering jet axes that preferentially align themselves with the heavy quarks themselves, we were able to reveal the dead-cone angle from exceedingly simple observables, namely the first two cumulants of particular energy distributions within jets. We went further to define analogous observables differential not in angular scales, but rather, transverse momentum

scales, which allow for the measurement of the dead-cone effect that is universal across collisional systems and center-of-mass energies.

Lastly, we developed a scheme for the tagging of top jets from a background of QCD jets that leveraged the natural metric space structure inherent to the point-cloud data type, does so in a completely unsupervised fashion, and relied only on a few parameters whose values are both physically intuitive and can be readily extracted from the data set itself. Furthermore, we developed a new data type for the representation of jets whose properties not only include the ability of embedding and visualizing ensembles of jets in a two-dimensional plane, but is also able to extract the characteristic angular size of the top decay products. The data type naturally picks out this angle as the unique internal resolution scale over which QCD and top jets are most readily distinguishable.

Though the approaches of each section differed substantially at face value, they may all be understood to fall under the umbrella of statistics. All of the tools of a modern particle physicist can ultimately be understood as particular instances of statistics, such as Monte-Carlo simulations, quantum-field theoretic calculations, and machine learning applications. This perhaps to be expected as an ideal framework from which to view jet physics, since after all, jets are highly-complex many-body phenomena. The main theme advanced in this work is that taking the statistical point of view affords the theorist and experimentalist alike a great deal of simplicity in observable design, as well as clarity in understanding.

In conclusion, we hope that the methods laid forth prove useful to the jet physics community and help to elucidate the dynamics inherent to quarks of heavy flavor. Kenneth Wilson taught the world how to organize our view of the laws of physics according to energy scales. With this in mind, we may consider it quite fortunate for QCD to give rise to jets, as jets provide an incredibly rich microscope through which we can view the workings of nature at energy scales unimaginable by the human mind.

[AAA17] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al. "Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector." *JHEP*, **12**:024, 2017.

[Aad10] Georges Aad et al. "Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.77$ TeV with the ATLAS Detector at the LHC." *Phys. Rev. Lett.*, **105**:252303, 2010.

[Aad11] Georges Aad et al. "Measurement of the inclusive and dijet cross-sections of b^- jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector." *Eur. Phys. J.*, **C71**:1846, 2011.

[Aad23] Georges Aad et al. "A precise determination of the strong-coupling constant from the recoil of Z bosons with the ATLAS experiment at $\sqrt{s} = 8$ TeV." 9 2023.

[Abe07] B. I. Abelev et al. "Energy dependence of pi+-, p and anti-p transverse momentum spectra for Au+Au collisions at s(NN)**(1/2) = 62.4 and 200-GeV." *Phys. Lett.*, **B655**:104–113, 2007.

[Acc16] A. Accardi et al. "Electron Ion Collider: The Next QCD Frontier." *Eur. Phys. J.*, **A52**(9):268, 2016.

[Ach18] S. Acharya et al. "First measurement of jet mass in PbPb and pPb collisions at the LHC." *Phys. Lett.*, **B776**:249–264, 2018.

[Ach22] S. Acharya et al. "Direct observation of the dead-cone effect in quantum chromodynamics." *Nature*, **605**(7910):440–446, 2022. [Erratum: Nature 607, E22 (2022)].

[ACM17] J. A. Aguilar-Saavedra, Jack H. Collins, and Rashmish K. Mishra. "A generic anti-QCD jet tagger." *JHEP*, **11**:163, 2017.

[Ada08] A. Adare et al. "Onset of pi0 Suppression Studied in Cu+Cu Collisions at sNN=22.4, 62.4, and 200 GeV." *Phys. Rev. Lett.*, **101**:162301, 2008.

[Ada15] A. Adare et al. "An Upgrade Proposal from the PHENIX Collaboration." 2015.

[Ada17] L. Adamczyk et al. "Dijet imbalance measurements in $Au + Au$ and pp collisions at $\sqrt{s_{NN}} = 200$GeV at STAR." *Phys. Rev. Lett.*, **119**(6):062301, 2017.

[ADM16] E.C. Aschenauer, U. D'Alesio, and F. Murgia. "TMDs and SSAs in hadronic interactions." *Eur. Phys. J. A*, **52**(6):156, 2016.

[Ado17] C. Adolph et al. "First measurement of the Sivers asymmetry for gluons using SIDIS data." *Phys. Lett. B*, **772**:854–864, 2017.

[AFF14] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." *JHEP*, **07**:079, 2014.

[AG04] Azfar Adil and Miklos Gyulassy. "Energy systematics of jet tomography at RHIC: s**(1/2) = 62.4 vs. 200-A-Gev." *Phys. Lett.*, **B602**:52–59, 2004.

[AGH07] A. Adil, M. Gyulassy, W. A. Horowitz, and S. Wicks. "Collisional Energy Loss of Non Asymptotic Jets in a QGP." *Phys. Rev.*, **C75**:044906, 2007.

[Air10] A. Airapetian et al. "Leading-Order Determination of the Gluon Polarization from high-p_T Hadron Electroproduction." *JHEP*, **08**:130, 2010.

[AKP20] Miguel Arratia, Zhong-Bo Kang, Alexei Prokudin, and Felix Ringer. "Jet-based measurements of Sivers and Collins asymmetries at the future electron-ion collider." *Phys. Rev. D*, **102**(7):074015, 2020.

[Alw08] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions." *Eur. Phys. J. C*, **53**:473–500, 2008.

[AMY02] Peter Brockway Arnold, Guy D. Moore, and Laurence G. Yaffe. "Photon and gluon emission in relativistic plasmas." *JHEP*, **06**:030, 2002.

[AP77] G. Altarelli and G. Parisi. "ASYMPTOTIC FREEDOM IN PARTON LANGUAGE." *Nucl. Phys.*, **B126**:298, 1977.

[Apr15] Ani Aprahamian et al. "Reaching for the horizon: The 2015 long range plan for nuclear science." 10 2015.

[Arb20] A. Arbuzov et al. "On the physics potential to study the gluon content of proton and deuteron at NICA SPD." 11 2020.

[ARV13] Daniele P. Anderle, Felix Ringer, and Werner Vogelsang. "QCD resummation for semi-inclusive hadron production processes." *Phys. Rev. D*, **87**(3):034014, 2013.

[BAO20] Alexander Bogatskiy, Brandon Anderson, Jan T. Offermann, Marwah Roussi, David W. Miller, and Risi Kondor. "Lorentz Group Equivariant Neural Network for Particle Physics.", 2020.

[BBC14] H. Berrehrah, E. Bratkovskaya, W. Cassing, P. B. Gossiaux, J. Aichelin, and M. Bleicher. "Collisional processes of on-shell and off-shell heavy quarks in vacuum and in the Quark-Gluon-Plasma." *Phys. Rev.*, **C89**(5):054901, 2014.

[BBE78] C. Louis Basham, Lowell S. Brown, Stephen D. Ellis, and Sherwin T. Love. "Energy Correlations in electron - Positron Annihilation: Testing QCD." *Phys. Rev. Lett.*, **41**:1585, 1978.

[BBF20] Marcel Balsiger, Thomas Becher, and Andrea Ferroglia. "Resummation of non-global logarithms in cross sections with massive particles." *JHEP*, **09**:029, 2020.

[BBP20] Alessandro Bacchetta, Daniël Boer, Cristian Pisano, and Pieter Taels. "Gluon TMDs and NRQCD matrix elements in J/ψ production at an EIC." *Eur. Phys. J. C*, **80**(1):72, 2020.

[BCT14a] Daniele Bertolini, Tucker Chan, and Jesse Thaler. "Jet Observables Without Jet Algorithms." *JHEP*, **04**:013, 2014.

[BCT14b] Daniele Bertolini, Tucker Chan, and Jesse Thaler. "Jet Observables Without Jet Algorithms." *JHEP*, **04**:013, 2014.

[BDD08] Andrea Banfi, Mrinal Dasgupta, and Yazid Delenda. "Azimuthal decorrelations between QCD jets at all orders." *Phys. Lett. B*, **665**:86–91, 2008.

[BDF22] Thorsten Buss, Barry M. Dillon, Thorben Finke, Michael Krämer, Alessandro Morandini, Alexander Mück, Ivan Oleksiyuk, and Tilman Plehn. "What's Anomalous in LHC Jets?" 2 2022.

[BDM97] R. Baier, Yuri L. Dokshitzer, Alfred H. Mueller, S. Peigne, and D. Schiff. "Radiative energy loss of high-energy quarks and gluons in a finite volume quark - gluon plasma." *Nucl. Phys.*, **B483**:291–320, 1997.

[BDM20] Daniël Boer, Umberto D'Alesio, Francesco Murgia, Cristian Pisano, and Pieter Taels. "J/ψ meson production in SIDIS: matching high and low transverse momentum." *JHEP*, **09**:040, 2020.

[Beu23] Jyotiranjan Beuria. "Intrinsic geometry of collider events and nearest neighbour based weighted filtration." 11 2023.

[BHM22] Alexander Bogatskiy, Timothy Hoffman, David W. Miller, and Jan T. Offermann. "PELICAN: Permutation Equivariant and Lorentz Invariant or Covariant Aggregator Network for Particle Physics." 11 2022.

[BHM23] Alexander Bogatskiy, Timothy Hoffman, David W. Miller, Jan T. Offermann, and Xiaoyang Liu. "Explainable Equivariant Neural Networks for Particle Physics: PELICAN.", 2023.

[BHS02] Stanley J. Brodsky, Dae Sung Hwang, and Ivan Schmidt. "Final state interactions and single spin asymmetries in semiinclusive deep inelastic scattering." *Phys. Lett. B*, **530**:99–107, 2002.

[Bie22] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3." 3 2022.

[BKL18] Maarten G. A. Buffing, Zhong-Bo Kang, Kyle Lee, and Xiaohui Liu. "A transverse momentum dependent framework for back-to-back photon+jet production." 2018.

[BKT10] Hiroo Beppu, Yuji Koike, Kazuhiro Tanaka, and Shinsuke Yoshida. "Contribution of Twist-3 Multi-Gluon Correlation Functions to Single Spin Asymmetry in Semi-Inclusive Deep Inelastic Scattering." *Phys. Rev. D*, **82**:054005, 2010.

[BLP15] Daniël Boer, Cédric Lorcé, Cristian Pisano, and Jian Zhou. "The gluon Sivers distribution: status and future prospects." *Adv. High Energy Phys.*, **2015**:371396, 2015.

[BM08] C.J. Bomhof and Piet J. Mulders. "Non-universality of transverse momentum dependent parton distribution functions." *Nucl. Phys. B*, **795**:409–427, 2008.

[BMM13] M.G.A. Buffing, A. Mukherjee, and P.J. Mulders. "Generalized Universality of Definite Rank Gluon Transverse Momentum Dependent Correlators." *Phys. Rev. D*, **88**:054027, 2013.

[BMP03] Daniel Boer, P.J. Mulders, and F. Pijlman. "Universality of T odd effects in single spin and azimuthal asymmetries." *Nucl. Phys. B*, **667**:201–241, 2003.

[BMP16] Danil Boer, Piet J. Mulders, Cristian Pisano, and Jian Zhou. "Asymmetries in Heavy Quark Pair and Dijet Production at an EIC." *JHEP*, **08**:001, 2016.

[BMP20] Alejandro Bris, Vicent Mateu, and Moritz Preisser. "Massive event-shape distributions at N^2LL." *JHEP*, **09**:132, 2020.

[Boe11] Daniel Boer et al. "Gluons and the quark sea at high energies: Distributions, polarization, tomography." 8 2011.

[BT91] Eric Braaten and Markus H. Thoma. "Energy loss of a heavy quark in the quark - gluon plasma." *Phys. Rev.*, **D44**(9):R2625, 1991.

[Bur04] Matthias Burkardt. "Sivers mechanism for gluons." *Phys. Rev.*, **D69**:091501, 2004.

[Bur14] Karen M. Burke et al. "Extracting the jet transport coefficient from jet quenching in high-energy heavy-ion collisions." *Phys. Rev.*, **C90**(1):014909, 2014.

[But19] Anja Butter et al. "The Machine Learning landscape of top taggers." *SciPost Phys.*, **7**:014, 2019.

[Cam03] Francesco Camastra. "Data dimensionality estimation methods: a survey." *Pattern Recognition*, **36**(12):2945–2954, 2003.

[Cao18] Shanshan Cao et al. "Towards the extraction of heavy-quark transport coefficients in quark-gluon plasma." 2018.

[CCC20] Tianji Cai, Junyi Cheng, Nathaniel Craig, and Katy Craig. "Linearized optimal transport for collider events." *Phys. Rev. D*, **102**(11):116019, 2020.

[CCC22] Tianji Cai, Junyi Cheng, Katy Craig, and Nathaniel Craig. "Which metric on the space of collider events?" *Phys. Rev. D*, **105**(7):076003, 2022.

[CEK16] Yang-Ting Chien, Alexander Emerman, Zhong-Bo Kang, Grigory Ovanesyan, and Ivan Vitev. "Jet Quenching from QCD Evolution." *Phys. Rev.*, **D93**(7):074030, 2016.

[CEM20] Rafael F. del Castillo, Miguel G. Echevarria, Yiannis Makris, and Ignazio Scimemi. "TMD factorization for di-jet and heavy meson pair in DIS." 8 2020.

[CG12] S. Catani and M. Grazzini. "Higgs Boson Production at Hadron Colliders: Hard-Collinear Coefficients at the NNLO." *Eur. Phys. J. C*, **72**:2013, 2012. [Erratum: Eur.Phys.J.C 72, 2132 (2012)].

[Cha11] Serguei Chatrchyan et al. "Observation and studies of jet quenching in PbPb collisions at nucleon-nucleon center-of-mass energy = 2.76 TeV." *Phys. Rev.*, **C84**:024906, 2011.

[Cha12] Serguei Chatrchyan et al. "Shape, Transverse Size, and Charged Hadron Multiplicity of Jets in pp Collisions at 7 TeV." *JHEP*, **06**:160, 2012.

[Cha13] Serguei Chatrchyan et al. "Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector." *Phys. Rev.*, **D87**(11):112002, 2013. [Erratum: Phys. Rev.D87,no.11,119902(2013)].

[Cha14] Serguei Chatrchyan et al. "Evidence of b-Jet Quenching in PbPb Collisions at $\sqrt{s_{NN}} = 2.76$TeV." *Phys. Rev. Lett.*, **113**(13):132301, 2014. [Erratum: Phys. Rev. Lett.115,no.2,029903(2015)].

[Che22] Yi Chen et al. "Jet energy spectrum and substructure in e^+e collisions at 91.2 GeV with ALEPH Archived Data." *JHEP*, **06**:008, 2022.

[CJN12] Jui-Yu Chiu, Ambar Jain, Duff Neill, and Ira Z. Rothstein. "A Formalism for the Systematic Treatment of Rapidity Logarithms in Quantum Field Theory." *JHEP*, **05**:084, 2012.

[CK18] Yang-Ting Chien and Raghav Kunnawalkam Elayavalli. "Probing heavy ion collisions using quark and gluon jet substructure." 3 2018.

[CKS15] Josh Cogan, Michael Kagan, Emanuel Strauss, and Ariel Schwartzman. "Jet-Images: Computer Vision Inspired Techniques for Jet Tagging." *JHEP*, **02**:118, 2015.

[CLM21] Yi Chen, Yen-Jie Lee, Marcello Maggi, Paoti Chang, Yang-Ting Chien, Christopher McGinn, and Dennis Perepelitsa. "Analysis note: jet reconstruction, energy spectra, and substructure analyses with archived ALEPH data." 8 2021.

[CLM22a] Simone Caletti, Andrew J. Larkoski, Simone Marzani, and Daniel Reichelt. "A fragmentation approach to jet flavor." *JHEP*, **10**:158, 2022.

[CLM22b] Evan Craft, Kyle Lee, Bianka Mecaj, and Ian Moult. "Beautiful and Charming Energy Correlators." 10 2022.

[CM21] Frdric Chazal and Bertrand Michel. "An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists." 2021.

[CNR18] Megan Connors, Christine Nattrass, Rosi Reed, and Sevil Salur. "Jet measurements in heavy ion physics." *Rev. Mod. Phys.*, **90**:025005, 2018.

[Col02] John C. Collins. "Leading twist single transverse-spin asymmetries: Drell-Yan and deep inelastic scattering." *Phys. Lett. B*, **536**:43–48, 2002.

[Col13] John Collins. *Foundations of perturbative QCD*, volume 32. Cambridge University Press, 11 2013.

[Col20] CMS Collaboration. "Machining Jets.", 2020.

[Col21] Jack H. Collins. "An Exploration of Learnt Representations of W Jets." 9 2021.

[Com23] Nuclear Science Advisory Committee. "A New Era of Discovery: The 2023 Long Range Plan For Nuclear Science.", 2023.

[CP19] Leticia Cunqueiro and Mateusz Ploskon. "Searching for the dead cone effects with iterative declustering of heavy-flavor jets." *Phys. Rev. D*, **99**(7):074027, 2019.

[CRW19a] Pedro Cal, Felix Ringer, and Wouter J. Waalewijn. "The Jet Shape at NLL′." *JHEP*, **05**:143, 2019.

[CRW19b] Pedro Cal, Felix Ringer, and Wouter J. Waalewijn. "The jet shape at NLL′." *JHEP*, **05**:143, 2019.

[CSS85] John C. Collins, Davison E. Soper, and George F. Sterman. "Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production." *Nucl. Phys.*, **B250**:199–224, 1985.

[CSS08] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k_t jet clustering algorithm." *JHEP*, **04**:063, 2008.

[CSS12] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual." *Eur. Phys. J. C*, **72**:1896, 2012.

[CSW19] Yang-Ting Chien, Ding Yu Shao, and Bin Wu. "Resummation of Boson-Jet Correlation at Hadron Colliders." *JHEP*, **11**:025, 2019.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.

[CV14a] Yang-Ting Chien and Ivan Vitev. "Jet Shape Resummation Using Soft-Collinear Effective Theory." *JHEP*, **12**:061, 2014.

[CV14b] Yang-Ting Chien and Ivan Vitev. "Jet Shape Resummation Using Soft-Collinear Effective Theory." *JHEP*, **12**:061, 2014.

[CV16a] Yang-Ting Chien and Ivan Vitev. "Towards the understanding of jet shapes and cross sections in heavy ion collisions using soft-collinear effective theory." *JHEP*, **05**:023, 2016.

[CV16b] Yang-Ting Chien and Ivan Vitev. "Towards the understanding of jet shapes and cross sections in heavy ion collisions using soft-collinear effective theory." *JHEP*, **05**:023, 2016.

[CV17] Yang-Ting Chien and Ivan Vitev. "Probing the Hardest Branching within Jets in Heavy-Ion Collisions." *Phys. Rev. Lett.*, **119**(11):112301, 2017.

[DDS16] Mrinal Dasgupta, Frdric A. Dreyer, Gavin P. Salam, and Gregory Soyez. "Inclusive jet spectrum for small-radius jets." *JHEP*, **06**:057, 2016.

[DFK20] B. M. Dillon, D. A. Faroughy, J. F. Kamenik, and M. Szewc. "Learning the latent structure of collider events." *JHEP*, **10**:206, 2020.

[DFM19] Umberto D'Alesio, Carlo Flore, Francesco Murgia, Cristian Pisano, and Pieter Taels. "Unraveling the Gluon Sivers Function in Hadronic Collisions at RHIC." *Phys. Rev. D*, **99**(3):036013, 2019.

[DGV06] Magdalena Djordjevic, Miklos Gyulassy, Ramona Vogt, and Simon Wicks. "Influence of bottom quark jet quenching on single electron tomography of Au + Au." *Phys. Lett. B*, **632**:81–86, 2006.

[DHG16] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. "New parton distribution functions from a global analysis of quantum chromodynamics." *Phys. Rev.*, **D93**(3):033006, 2016.

[Die17] Reinhard Diestel. *Graph Theory*. Springer Berlin, Heidelberg, GR, 2017.

[DK01] Yuri L. Dokshitzer and D. E. Kharzeev. "Heavy quark colorimetry of QCD matter." *Phys. Lett.*, **B519**:199–206, 2001.

[DKL16] Lin Dai, Chul Kim, and Adam K. Leibovich. "Fragmentation of a Jet with Small Radius." *Phys. Rev.*, **D94**(11):114023, 2016.

[DKL18] Lin Dai, Chul Kim, and Adam K. Leibovich. "Heavy Quark Jet Fragmentation." *JHEP*, **09**:109, 2018.

[DKL21] Lin Dai, Chul Kim, and Adam K. Leibovich. "Heavy quark jet production near threshold." *JHEP*, **09**:148, 2021.

[DKP15] Ling-Yun Dai, Zhong-Bo Kang, Alexei Prokudin, and Ivan Vitev. "Next-to-leading order transverse momentum-weighted Sivers asymmetry in semi-inclusive deep inelastic scattering: the role of the three-gluon correlator." *Phys. Rev. D*, **92**(11):114024, 2015.

[DKT91] Yuri L. Dokshitzer, Valery A. Khoze, and S. I. Troian. "On specific QCD properties of heavy quark fragmentation ('dead cone')." *J. Phys. G*, **17**:1602–1604, 1991.

[DMP15] U. D'Alesio, F. Murgia, and C. Pisano. "Towards a first estimate of the gluon Sivers function from A_N data in pp collisions at RHIC." *JHEP*, **09**:119, 2015.

[DMP19] Umberto D'Alesio, Francesco Murgia, Cristian Pisano, and Pieter Taels. "Azimuthal asymmetries in semi-inclusive J/ψ + jet production at an EIC." *Phys. Rev. D*, **100**(9):094016, 2019.

[DPS21] Barry M. Dillon, Tilman Plehn, Christof Sauer, and Peter Sorrenson. "Better Latent Spaces for Better Autoencoders." *SciPost Phys.*, **11**:061, 2021.

[DS01] M. Dasgupta and G. P. Salam. "Resummation of nonglobal QCD observables." *Phys. Lett.*, **B512**:323–330, 2001.

[DWZ18a] Wei Dai, Sa Wang, Shan-Liang Zhang, Ben-Wei Zhang, and Enke Wang. "Transverse Momentum Balance and Angular Distribution of $b\bar{b}$ Dijets in Pb+Pb collisions." 2018.

[DWZ18b] Wei Dai, Sa Wang, Shan-Liang Zhang, Ben-Wei Zhang, and Enke Wang. "Transverse Momentum Balance and Angular Distribution of $b\bar{b}$ Dijets in Pb+Pb collisions." 2018.

[EIK14] Miguel G. Echevarria, Ahmad Idilbi, Zhong-Bo Kang, and Ivan Vitev. "QCD Evolution of the Sivers Asymmetry." *Phys. Rev.*, **D89**:074013, 2014.

[EKM15] Miguel G. Echevarria, Tomas Kasemets, Piet J. Mulders, and Cristian Pisano. "QCD evolution of (un)polarized gluon TMDPDFs and the Higgs q_T-distribution." *JHEP*, **07**:158, 2015. [Erratum: JHEP 05, 073 (2017)].

[EKS96] M. Ester, H.P. Kriegel, J. Sander, and X. Xu. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, 1996.

[ELZ03] V. V. Ezhela, S. B. Lugovsky, and O. V. Zenin. "Hadronic part of the muon g-2 estimated on the sigma**2003(tot)(e+ e- —> hadrons) evaluated data compilation." 12 2003.

[EMV20] Markus A. Ebert, Bernhard Mistlberger, and Gherardo Vita. "Transverse momentum dependent PDFs at N^3LO." 6 2020.

[ES93] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions." *Phys. Rev. D*, **48**:3160–3166, 1993.

[ESV16] Miguel G. Echevarria, Ignazio Scimemi, and Alexey Vladimirov. "Unpolarized Transverse Momentum Dependent Parton Distribution and Fragmentation Functions at next-to-next-to-leading order." *JHEP*, **09**:004, 2016.

[ESW11] R. Keith Ellis, W. James Stirling, and B. R. Webber. *QCD and collider physics*, volume 8. Cambridge University Press, 2 2011.

[ESZ19] Markus A. Ebert, Iain W. Stewart, and Yong Zhao. "Towards Quasi-Transverse Momentum Dependent PDFs Computable on the Lattice." *JHEP*, **09**:037, 2019.

[EVW10] Stephen D. Ellis, Christopher K. Vermilion, Jonathan R. Walsh, Andrew Hornig, and Christopher Lee. "Jet Shapes and Jet Algorithms in SCET." *JHEP*, **11**:101, 2010.

[FDD14] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. "DELPHES 3, A modular framework for fast simulation of a generic collider experiment." *JHEP*, **02**:057, 2014.

[FHM08a] Sean Fleming, Andre H. Hoang, Sonny Mantry, and Iain W. Stewart. "Jets from massive unstable particles: Top-mass determination." *Phys. Rev. D*, **77**:074010, 2008.

[FHM08b] Sean Fleming, Andre H. Hoang, Sonny Mantry, and Iain W. Stewart. "Top Jets in the Peak Region: Factorization Analysis with NLL Resummation." *Phys. Rev. D*, **77**:114003, 2008.

[FHM22] Katherine Fraser, Samuel Homiller, Rashmish K. Mishra, Bryan Ostdiek, and Matthew D. Schwartz. "Challenges for unsupervised anomaly detection in particle physics." *JHEP*, **03**:066, 2022.

[FN21] Matthew Feickert and Benjamin Nachman. "A Living Review of Machine Learning for Particle Physics." 2 2021.

[FTW21] Taylor Faucett, Jesse Thaler, and Daniel Whiteson. "Mapping Machine-Learned Physics into a Human-Readable Space." *Phys. Rev. D*, **103**(3):036020, 2021.

[GLV00] M. Gyulassy, P. Levai, and I. Vitev. "NonAbelian energy loss at finite opacity." *Phys. Rev. Lett.*, **85**:5535–5538, 2000.

[GLY14] Thomas Gehrmann, Thomas Luebbert, and Li Lin Yang. "Calculation of the transverse parton distribution functions at next-to-next-to-leading order." *JHEP*, **06**:155, 2014.

[GP83] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors." *Phys. Rev. Lett.*, **50**:346–349, Jan 1983.

[Gro22] Franz Gross et al. "50 Years of Quantum Chromodynamics." 12 2022.

[GSW18] Daniel Gutierrez-Reyes, Ignazio Scimemi, Wouter J. Waalewijn, and Lorenzo Zoppi. "Transverse momentum dependent distributions with jets." *Phys. Rev. Lett.*, **121**(16):162001, 2018.

[GSW19] Daniel Gutierrez-Reyes, Ignazio Scimemi, Wouter J. Waalewijn, and Lorenzo Zoppi. "Transverse momentum dependent distributions in e^+e^- and semi-inclusive deep-inelastic scattering using jets." *JHEP*, **10**:031, 2019.

[GW73] D.J. Gross and F. Wilczek. "ULTRAVIOLET BEHAVIOR OF NON-ABELIAN GAUGE THEORIES." *Phys. Rev. Lett.*, **30**:1343–1346, 1973.

[HDP22] Greg Hamilton, Travis Dore, and Christopher Plumberg. "Applications of persistent homology in nuclear collisions." *Phys. Rev. C*, **106**(6):064912, 2022.

[HKP19] Theo Heimel, Gregor Kasieczka, Tilman Plehn, and Jennifer M. Thompson. "QCD or What?" *SciPost Phys.*, **6**(3):030, 2019.

[HKV13a] Jinrui Huang, Zhong-Bo Kang, and Ivan Vitev. "Inclusive b-jet production in heavy ion collisions at the LHC." *Phys. Lett.*, **B726**:251–256, 2013.

[HKV13b] Jinrui Huang, Zhong-Bo Kang, and Ivan Vitev. "Inclusive b-jet production in heavy ion collisions at the LHC." *Phys. Lett. B*, **726**:251, 2013.

[HKV15a] Jinrui Huang, Zhong-Bo Kang, Ivan Vitev, and Hongxi Xing. "Photon-tagged and B-meson-tagged b-jet production at the LHC." *Phys. Lett.*, **B750**:287–293, 2015.

[HKV15b] Jinrui Huang, Zhong-Bo Kang, Ivan Vitev, and Hongxi Xing. "Photon-tagged and B-meson-tagged b-jet production at the LHC." *Phys. Lett. B*, **750**:287, 2015.

[HLS19] André H. Hoang, Christopher Lepenik, and Maximilian Stahlhofen. "Two-Loop Massive Quark Jet Functions in SCET." *JHEP*, **08**:112, 2019.

[HVZ12] Yuncun He, Ivan Vitev, and Ben-Wei Zhang. "$\mathcal{O}(\alpha_s^3)$ Analysis of Inclusive Jet and di-Jet Production in Heavy Ion Reactions at the Large Hadron Collider." *Phys. Lett.*, **B713**:224–232, 2012.

[Ji92] Xiang-Dong Ji. "Gluon correlations in the transversely polarized nucleon." *Phys. Lett. B*, **289**:137–142, 1992.

[JL11] Martin Jankowiak and Andrew J. Larkoski. "Jet Substructure Without Trees." *JHEP*, **06**:057, 2011.

[JSV04] B. Jager, M. Stratmann, and W. Vogelsang. "Single inclusive jet production in polarized pp collisions at $O(\alpha_s^3)$." *Phys. Rev.*, **D70**:034010, 2004.

[K02] Balázs Kégl. "Intrinsic Dimension Estimation Using Packing Numbers." In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002.

[Kas21] Gregor Kasieczka et al. "The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics." *Rept. Prog. Phys.*, **84**(12):124201, 2021.

[Kha16] Vardan Khachatryan et al. "Measurement of transverse momentum relative to dijet systems in PbPb and pp collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV." *JHEP*, **01**:006, 2016.

[Kim20] Chul Kim. "Exclusive heavy quark dijet cross section." *J. Korean Phys. Soc.*, **77**(6):469–476, 2020.

[KKT22] Patrick T. Komiske, Serhii Kryhin, and Jesse Thaler. "Disentangling Quarks and Gluons with CMS Open Data." 5 2022.

[KLO15] Zhong-Bo Kang, Robin Lashof-Regas, Grigory Ovanesyan, Philip Saad, and Ivan Vitev. "Jet quenching phenomenology from soft-collinear effective theory with Glauber gluons." *Phys. Rev. Lett.*, **114**(9):092002, 2015.

[KLS20] Zhong-Bo Kang, Kyle Lee, Ding Yu Shao, and John Terry. "The Sivers Asymmetry in Hadronic Dijet Production." 8 2020.

[KMM20] Patrick T. Komiske, Radha Mastandrea, Eric M. Metodiev, Preksha Naik, and Jesse Thaler. "Exploring the Space of Jets with CMS Open Data." *Phys. Rev. D*, **101**(3):034009, 2020.

[KMS17] Patrick T. Komiske, Eric M. Metodiev, and Matthew D. Schwartz. "Deep learning in color: towards automated quark/gluon jet discrimination." *JHEP*, **01**:110, 2017.

[KMT19a] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. "Energy Flow Networks: Deep Sets for Particle Jets." *JHEP*, **01**:121, 2019.

[KMT19b] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. "Energy flow networks: deep sets for particle jets." *Journal of High Energy Physics*, **2019**(1), 2019.

[KMT19c] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. "Metric Space of Collider Events." *Phys. Rev. Lett.*, **123**(4):041801, 2019.

[KMT20] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. "The Hidden Geometry of Particle Collisions." *JHEP*, **07**:006, 2020.

[Kog19] Roman Kogler et al. "Jet Substructure at the Large Hadron Collider: Experimental Review." *Rev. Mod. Phys.*, **91**(4):045003, 2019.

[KPR17] Gregor Kasieczka, Tilman Plehn, Michael Russell, and Torben Schell. "Deep-learning Top Taggers or The End of QCD?" *JHEP*, **05**:006, 2017.

[Kre00] S. Kretzer. "Fragmentation functions from flavour-inclusive and flavour- tagged e+ e- annihilations." *Phys. Rev.*, **D62**:054001, 2000.

[KRS21] Zhong-Bo Kang, Jared Reiten, Ding Yu Shao, and John Terry. "QCD evolution of the gluon Sivers function in heavy flavor dijet production at the Electron-Ion Collider." *JHEP*, **05**:286, 2021.

[KRV16] Zhong-Bo Kang, Felix Ringer, and Ivan Vitev. "The semi-inclusive jet function in SCET and small radius resummation for inclusive jet production." *JHEP*, **10**:125, 2016.

[KRV17a] Zhong-Bo Kang, Felix Ringer, and Ivan Vitev. "Effective field theory approach to open heavy flavor production in heavy-ion collisions." *JHEP*, **03**:146, 2017.

[KRV17b] Zhong-Bo Kang, Felix Ringer, and Ivan Vitev. "Inclusive production of small radius jets in heavy-ion collisions." *Phys. Lett.*, **B769**:242–248, 2017.

[KRV19a] Zhong-Bo Kang, Jared Reiten, Ivan Vitev, and Boram Yoon. "Light and heavy flavor dijet production and dijet mass modification in heavy ion collisions." *Phys. Rev. D*, **99**(3):034006, 2019.

[KRV19b] Zhong-Bo Kang, Jared Reiten, Ivan Vitev, and Boram Yoon. "Light and heavy flavor dijet production and dijet mass modification in heavy ion collisions." *Phys. Rev. D*, **99**:034006, 2019.

[KRW17] Zhong-Bo Kang, Felix Ringer, and Wouter J. Waalewijn. "The Energy Distribution of Subjets and the Jet Shape." *JHEP*, **07**:064, 2017.

[KTW10] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jet Trimming." *JHEP*, **02**:084, 2010.

[KVX17] Zhong-Bo Kang, Ivan Vitev, and Hongxi Xing. "Vector-boson-tagged jet production in heavy ion collisions at energies available at the CERN Large Hadron Collider." *Phys. Rev.*, **C96**(1):014912, 2017.

[LLX20] Lingfeng Li, Tao Liu, and Si-Jun Xu. "Jet Topology." 6 2020.

[LLY11] Hsiang-nan Li, Zhao Li, and C. P. Yuan. "QCD resummation for jet substructures." *Phys. Rev. Lett.*, **107**:152001, 2011.

[LLY13] Hsiang-nan Li, Zhao Li, and C. P. Yuan. "QCD resummation for light-particle jets." *Phys. Rev. D*, **87**:074025, 2013.

[LM20] Christopher Lepenik and Vicent Mateu. "NLO Massive Event-Shape Differential and Cumulative Distributions." *JHEP*, **03**:024, 2020.

[LMN17] Andrew J. Larkoski, Ian Moult, and Benjamin Nachman. "Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning." 2017.

[LMR18] Xiaohui Liu, Sven-Olaf Moch, and Felix Ringer. "Phenomenology of single-inclusive jet production with jet radius and threshold resummation." *Phys. Rev. D*, **97**(5):056026, 2018.

[LN22] Sung Hak Lim and Mihoko M. Nojiri. "Morphology for jet classification." *Phys. Rev. D*, **105**(1):014004, 2022.

[LN23] Andrew J. Larkoski and Duff Neill. "Flavor Fragmentation Function Factorization." 10 2023.

[LNT14] Andrew J. Larkoski, Duff Neill, and Jesse Thaler. "Jet Shapes with the Broadening Axis." *JHEP*, **04**:017, 2014.

[LOS16] Thomas Lübbert, Joel Oredsson, and Maximilian Stahlhofen. "Rapidity renormalized TMD soft and beam functions at two loops." *JHEP*, **03**:168, 2016.

[LRV19] Xiaohui Liu, Felix Ringer, Werner Vogelsang, and Feng Yuan. "Lepton-jet Correlations in Deep Inelastic Scattering at the Electron-Ion Collider." *Phys. Rev. Lett.*, **122**(19):192003, 2019.

[LRV20] Xiaohui Liu, Felix Ringer, Werner Vogelsang, and Feng Yuan. "Lepton-jet Correlation in Deep Inelastic Scattering." 7 2020.

[LST13] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. "Energy Correlation Functions for Jet Substructure." *JHEP*, **06**:108, 2013.

[LT23] Andrew J. Larkoski and Jesse Thaler. "A Spectral Metric for Collider Geometry." 5 2023.

[LV19a] Hai Tao Li and Ivan Vitev. "Inclusive heavy flavor jet production with semi-inclusive jet functions: from proton to heavy-ion collisions." *JHEP*, **07**:148, 2019.

[LV19b] Hai Tao Li and Ivan Vitev. "Inclusive heavy flavor jet production with semi-inclusive jet functions: from proton to heavy-ion collisions." *JHEP*, **07**:148, 2019.

[LV19c] Hai Tao Li and Ivan Vitev. "Inverting the mass hierarchy of jet quenching effects with prompt b-jet substructure." *Phys. Lett. B*, **793**:259, 2019.

[LYZ20] Ming-Xing Luo, Tong-Zhi Yang, Hua Xing Zhu, and Yu Jiao Zhu. "Transverse Parton Distribution and Fragmentation Functions at NNLO: the Gluon Case." *JHEP*, **01**:040, 2020.

[MC21] Vinicius Mikuni and Florencia Canelli. "Unsupervised clustering for collider physics." *Phys. Rev. D*, **103**(9):092007, 2021.

[MCD20] Eric A. Moreno, Olmo Cerri, Javier M. Duarte, Harvey B. Newman, Thong Q. Nguyen, Avikar Periwal, Maurizio Pierini, Aidana Serikova, Maria Spiropulu, and Jean-Roch Vlimant. "JEDI-net: a jet identification algorithm based on interaction networks." *Eur. Phys. J. C*, **80**(1):58, 2020.

[MHA17] Leland McInnes, John Healy, and Steve Astels. "hdbscan: Hierarchical density based clustering." *The Journal of Open Source Software*, **2**(11), mar 2017.

[MR01] P.J. Mulders and J. Rodrigues. "Transverse momentum dependence in gluon distribution and fragmentation functions." *Phys.Rev.*, **D63**:094021, 2001.

[MR17] Asmita Mukherjee and Sangem Rajesh. "J/ψ production in polarized and unpolarized ep collision and Sivers and $\cos 2\phi$ asymmetries." *Eur. Phys. J. C*, **77**(12):854, 2017.

[MRS07] Michael L. Miller, Klaus Reygers, Stephen J. Sanders, and Peter Steinberg. "Glauber modeling in high energy nuclear collisions." *Ann. Rev. Nucl. Part. Sci.*, **57**:205–243, 2007.

[MSS19] Simone Marzani, Gregory Soyez, and Michael Spannowsky. *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, volume 958. Springer, 2019.

[MST16] Fabio Maltoni, Michele Selvaggi, and Jesse Thaler. "Exposing the dead cone effect with jet substructure techniques." *Phys. Rev. D*, **94**(5):054015, 2016.

[MV12] Asmita Mukherjee and Werner Vogelsang. "Jet production in (un)polarized pp collisions: dependence on jet algorithm." *Phys. Rev.*, **D86**:094009, 2012.

[MV18] Yiannis Makris and Varun Vaidya. "Transverse Momentum Spectra at Threshold for Groomed Heavy Quark Jets." *JHEP*, **10**:019, 2018.

[NBC20] Ashwin Narayan, Bonnie Berger, and Hyunghoon Cho. "Density-Preserving Data Visualization Unveils Dynamic Patterns of Single-Cell Transcriptomic Variability." 2020.

[NLL19] Chien-Chun Ni, Yu-Yao Lin, Feng Luo, and Jie Gao. "Community Detection on Networks with Ricci Flow.", 2019.

[NPW19a] Duff Neill, Andreas Papaefstathiou, Wouter J. Waalewijn, and Lorenzo Zoppi. "Phenomenology with a recoil-free jet axis: TMD fragmentation and the jet shape." *JHEP*, **01**:067, 2019.

[NPW19b] Duff Neill, Andreas Papaefstathiou, Wouter J. Waalewijn, and Lorenzo Zoppi. "Phenomenology with a recoil-free jet axis: TMD fragmentation and the jet shape." *JHEP*, **01**:067, 2019.

[NR22] Mark S. Neubauer and Avik Roy. "Explainable AI for High Energy Physics." In *Snowmass 2021*, 6 2022.

[NSW17] Duff Neill, Ignazio Scimemi, and Wouter J. Waalewijn. "Jet axes and universal transverse-momentum-dependent fragmentation." *JHEP*, **04**:020, 2017.

[NVX14] R. B. Neufeld, Ivan Vitev, and Hongxi Xing. "Operator definition and derivation of collisional energy and momentum loss in relativistic plasmas." *Phys. Rev.*, **D89**(9):096003, 2014.

[OKM16] Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman. "Jet-images — deep learning edition." *JHEP*, **07**:069, 2016.

[OV12] Grigory Ovanesyan and Ivan Vitev. "Medium-induced parton splitting kernels from Soft Collinear Effective Theory with Glauber gluons." *Phys. Lett.*, **B706**:371–378, 2012.

[PBD22] Tilman Plehn, Anja Butter, Barry Dillon, and Claudius Krause. "Modern Machine Learning for LHC Physicists." 11 2022.

[PHO22] Sang Eon Park, Philip Harris, and Bryan Ostdiek. "Neural Embedding: Learning the Embedding of the Manifold of Physics Data." 8 2022.

[Pol73] H.D. Politzer. "RELIABLE PERTURBATIVE RESULTS FOR STRONG INTERACTIONS?" *Phys. Rev. Lett.*, **30**:1346–1349, 1973.

[PVG11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, **12**:2825–2830, 2011.

[QG20] Huilin Qu and Loukas Gouskos. "Jet tagging via particle clouds." *Physical Review D*, **101**(5), 2020.

[QM11] Guang-You Qin and Berndt Muller. "Explanation of Di-jet asymmetry in Pb+Pb collisions at the Large Hadron Collider." *Phys. Rev. Lett.*, **106**:162302, 2011. [Erratum: Phys. Rev. Lett.108,189904(2012)].

[QS91] Jian-wei Qiu and George F. Sterman. "Single transverse spin asymmetries." *Phys. Rev. Lett.*, **67**:2264–2267, 1991.

[QS92] Jian-wei Qiu and George F. Sterman. "Single transverse spin asymmetries in direct photon production." *Nucl. Phys. B*, **378**:52–78, 1992.

[RKM18] Sangem Rajesh, Raj Kishore, and Asmita Mukherjee. "Sivers effect in Inelastic J/ψ Photoproduction in $ep^{\uparrow}$ Collision in Color Octet Model." *Phys. Rev. D*, **98**(1):014007, 2018.

[RW23] Alexis Romero and Daniel Whiteson. "Jet Rotational Metrics." 11 2023.

[Sch14] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[Sch21] Matthew D. Schwartz. "Modern Machine Learning and Particle Physics." 3 2021.

[Sey97] Michael H. Seymour. "Jet Phenomenology." 1997.

[Sey98] Michael H. Seymour. "Jet Shapes in Hadron Collisions: Higher Orders, Resummation and Hadronization." *Nucl. Phys. B*, **513**:269, 1998.

[Sir18] Albert M Sirunyan et al. "Comparing transverse momentum balance of b jet pairs in pp and PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV." *JHEP*, **03**:181, 2018.

[Sir20] Albert M Sirunyan et al. "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques." *JINST*, **15**(06):P06005, 2020.

[SIY18] Peng Sun, Joshua Isaacson, C. P. Yuan, and Feng Yuan. "Nonperturbative functions for SIDIS and Drell–Yan processes." *Int. J. Mod. Phys. A*, **33**(11):1841006, 2018.

[SMG21] Tim Sainburg, Leland McInnes, and Timothy Q Gentner. "Parametric UMAP Embeddings for Representation and Semisupervised Learning." *Neural Computation*, **33**(11):2881–2907, 2021.

[SMJ16] R P Sreejith, Karthikeyan Mohanraj, Jrgen Jost, Emil Saucan, and Areejit Samal. "Forman curvature for complex networks." *Journal of Statistical Mechanics: Theory and Experiment*, **2016**(6):063206, June 2016.

[SMS06] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual." *JHEP*, **05**:026, 2006.

[SMS08] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1." *Comput. Phys. Commun.*, **178**:852–867, 2008.

[sPH] "sPHENIX: Heavy Flavor Jet Simulation and Analysis." https://www.phenix.bnl.gov/WWW/publish/jinhuang/sPHENIX/sPH-HF-2017-001-v1.1.pdf.

[SSE17] E. Schubert, J. Sander, M. Ester, H.P. Kriegel, and X. Xu. "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN." *ACM Transactions on Database Systems (TODS)*, **42(3)**(19), 2017.

[STA18] A.M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Ero, and et al. "Evidence for the Higgs boson decay to a bottom quark-antiquark pair." *Phys. Lett. B*, **780**:501, 2018.

[STV19] Ignazio Scimemi, Andrey Tarasov, and Alexey Vladimirov. "Collinear matching for Sivers function at next-to-leading order." *JHEP*, **05**:125, 2019.

[SW77] George F. Sterman and Steven Weinberg. "Jets from Quantum Chromodynamics." *Phys. Rev. Lett.*, **39**:1436, 1977.

[TDK22] Dawson Thomas, Sarah Demers, Smita Krishnaswamy, and Bastian Rieck. "Topological Jet Tagging." *NeurIPS*, 2022.

[Tho09] Markus H. Thoma. "Field Theoretic Description of Ultrarelativistic Electron-Positron Plasmas." *Rev. Mod. Phys.*, **81**:959–968, 2009.

[Tka97] Fyodor V. Tkachov. "Measuring multi - jet structure of hadronic energy flow or What is a jet?" *Int. J. Mod. Phys. A*, **12**:5411–5529, 1997.

[Vil09] Cedric Villani. *Optimal Transport: Old and New*. Springer-Verlag Berlin Heidelberg, GR, 2009.

[Vit05] Ivan Vitev. "Jet quenching at intermediate RHIC energies." *Phys. Lett.*, **B606**:303–312, 2005.

[Vit07] Ivan Vitev. "Non-Abelian energy loss in cold nuclear matter." *Phys. Rev.*, **C75**:064906, 2007.

[VWZ08a] Ivan Vitev, Simon Wicks, and Ben-Wei Zhang. "A Theory of jet shapes and cross sections: From hadrons to nuclei." *JHEP*, **11**:093, 2008.

[VWZ08b] Ivan Vitev, Simon Wicks, and Ben-Wei Zhang. "A theory of jet shapes and cross sections: from hadrons to nuclei." *JHEP*, **11**:093, 2008.

[Wei] Steven Weinberg. "The Quantum theory of fields. Vol. 1: Foundations." Cambridge, UK: Univ. Pr. (1995) 609 p.

[WG01] Xin-Nian Wang and Xiao-feng Guo. "Multiple parton scattering in nuclei: Parton energy loss." *Nucl. Phys.*, **A696**:788–832, 2001.

[WHD07] Simon Wicks, William Horowitz, Magdalena Djordjevic, and Miklos Gyulassy. "Elastic, inelastic, and path length fluctuations in jet tomography." *Nucl. Phys.*, **A784**:426–442, 2007.

[Wie00] Urs Achim Wiedemann. "Gluon radiation off hard quarks in a nuclear environment: Opacity expansion." *Nucl. Phys.*, **B588**:303–344, 2000.

[WW05] Qun Wang and Xin-nian Wang. "The Non-Abelian feature of parton energy loss in energy dependence of jet quenching in high-energy heavy-ion collisions." *Phys. Rev.*, **C71**:014903, 2005.

[XGD17] Saining Xie, Ross Girshick, Piotr Dollr, Zhuowen Tu, and Kaiming He. "Aggregated Residual Transformations for Deep Neural Networks." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, 2017.

[YSJ11] Clint Young, Bjorn Schenke, Sangyong Jeon, and Charles Gale. "Dijet asymmetry at the energies available at the CERN Large Hadron Collider." *Phys. Rev.*, **C84**:024907, 2011.

[Yua08] Feng Yuan. "Heavy Quarkonium Production in Single Transverse Polarized High Energy Scattering." *Phys. Rev. D*, **78**:014024, 2008.

[Zak97] B. G. Zakharov. "Radiative energy loss of high-energy quarks in finite size nuclear matter and quark - gluon plasma." *JETP Lett.*, **65**:615–620, 1997.

[ZAL18] Liang Zheng, E.C. Aschenauer, J.H. Lee, Bo-Wen Xiao, and Zhong-Bao Yin. "Accessing the gluon Sivers function at a future electron-ion collider." *Phys. Rev. D*, **98**(3):034011, 2018.

[Zar21] Nima Zardoshti. "First Direct Observation of the Dead-Cone Effect." *Nucl. Phys. A*, **1005**:121905, 2021.

[Zie95] Gunter M. Ziegler. *Lectures on Polytopes*. Springer New York, USA, 1995.

[ZWW04] Ben-Wei Zhang, Enke Wang, and Xin-Nian Wang. "Heavy quark energy loss in nuclear medium." *Phys. Rev. Lett.*, **93**:072301, 2004.

[Zyl20] P. A. Zyla et al. "Review of Particle Physics." *PTEP*, **2020**(8):083C01, 2020.