Understanding Jet Physics at Modern Particle Colliders

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citable link  http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050020

Terms of Use  This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Understanding Jet Physics at Modern Particle Colliders

A DISSERTATION PRESENTED
BY
CHRISTOPHER ROBERT FRYE
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
APRIL 2018
Understanding Jet Physics at Modern Particle Colliders

Abstract

We explore the phenomenon of jets at particle colliders from several diverse vantage points, beginning with a non-technical introduction to the field in general. Jets are messy objects that are difficult to model precisely; so much so, that the best way to design a high-precision observable is to avoid them. In the second chapter of this thesis, we do just that by introducing several electroweak observables that can be measured and computed to high precision, by carefully controlling the contribution of jets. In the third chapter, we attack the problem of jets at high precision head on by computing a jet substructure observable, the mass, to unprecedented accuracy. This is made possible through a grooming technique that removes the most complicated radiation from the jet before its mass is measured. In the fourth chapter, we make qualitative progress in understanding one of the most fundamental problems in jet physics: quark/gluon tagging. To do so, we introduce a new counting observable for jet substructure, and we glean insights through its quantitative calculation. In the final chapter, we explore jets from an entirely different angle, using techniques from modern machine learning. While such methods are generally opaque to interpretation, we build a framework for users to intuitively extract physics that the machine learns.
Contents

1 INTRODUCTION ........................................... 1
  1.1 Many Useful Theories .................................. 2
  1.2 Effective Field Theories in Particle Physics .......... 4
  1.3 Jets at Particle Colliders ................................ 7
  1.4 Overview of Dissertation ................................ 10

2 PRECISION DIBOSON OBSERVABLES FOR THE LHC .......... 13
  2.1 Introduction ........................................... 14
  2.2 Executive summary ...................................... 17
  2.3 The story at leading order ................................ 25
    2.3.1 High energy limit .................................. 25
    2.3.2 Squared amplitudes ................................ 27
    2.3.3 Partonic cross sections at high energies .......... 29
      2.3.3.1 $\gamma\gamma$, $Z\gamma$, $ZZ$ ................. 30
      2.3.3.2 $W^+\gamma$, $W^\pm Z$ .......................... 31
      2.3.3.3 $W^-W^+$ ....................................... 32
    2.3.4 Convolution with PDFs ................................ 34
    2.3.5 Ratio observables ................................... 38
    2.3.6 Limitations of finite statistics .................... 46
  2.4 Beyond leading order for $\gamma\gamma$, $Z\gamma$, $ZZ$ ....... 48
    2.4.1 Choices of observable and of cuts .................. 50
    2.4.2 NLO QCD corrections ................................ 53
    2.4.3 NNLO QCD corrections ................................ 58
    2.4.4 Partial cancellation of PDF and scale uncertainties 65
    2.4.5 EW corrections ...................................... 70
      2.4.5.1 Sudakov enhancements ........................... 70
      2.4.5.2 Proper choice of EW scales for on-shell external photons 74
  2.5 Additional practical considerations ..................... 76
    2.5.1 Photon isolation .................................... 76
    2.5.2 $Z$ decay and lepton isolation ....................... 80
  2.6 Discussion and summary .................................. 84
    2.6.1 Uncertainty budget .................................. 84
    2.6.2 Final comments ...................................... 87
      2.6.2.1 General reflections on our methods ............ 87

iv
2.6.2.2 Other Z decays .................................. 89
2.6.2.3 Applications of the $R_1$ ratios .............. 91
2.6.2.4 Prospects for the other ratios ................ 92

3 JET SUBSTRUCTURE BEYOND THE NEXT-TO-LEADING LOGARITHM 94
  3.1 Introduction ........................................ 95
  3.2 Observables ......................................... 102
    3.2.1 Soft Drop Grooming Algorithm .............. 102
    3.2.2 Energy Correlation Functions .............. 104
  3.3 Factorization Theorem ............................ 105
    3.3.1 Power Counting and Modes ................. 105
    3.3.2 Factorization and refactorization .......... 109
    3.3.3 The Single Scale of the Collinear-Soft Function .... 114
  3.4 Consequences of Factorization Theorem .......... 118
    3.4.1 Absence of Non-Global Logarithms to All Orders ... 119
    3.4.2 Process Independence .......................... 122
    3.4.3 Hadronization Corrections ................... 123
  3.5 Achieving NNLL Accuracy ........................... 125
    3.5.1 NNLL for $\alpha = 2$, $\beta = 0$ .......... 127
      3.5.1.1 Two-Loop Soft Function ............... 128
      3.5.1.2 Two-Loop Anomalous Dimension .......... 134
    3.5.2 Reclustering with anti-$k_T$ ............... 136
    3.5.3 NNLL for $\alpha = 2$, $\beta \geq 0$ .......... 140
  3.6 Matching NNLL to Fixed Order ..................... 144
    3.6.1 Matching Resummation to Fixed-Order ....... 145
    3.6.2 Comparison to Monte Carlo .................. 148
  3.7 Matching to Fixed Order in $pp \to Z + j$ ........ 155
    3.7.1 Resummed Cross Section in $pp \to Z + j$ .... 156
    3.7.2 Matching Resummation to Fixed-Order ........ 161
    3.7.3 Comparison to Monte Carlo .................. 165
  3.8 Conclusions ....................................... 169

4 OPTIMIZATION OF CALCULABLE QUARK/GLUON DISCRIMINANTS 172
  4.1 Introduction ....................................... 173
  4.2 Counting Observables with Soft Drop ............. 178
    4.2.1 Iterated Soft Drop ............................ 178
    4.2.2 Soft Drop Multiplicity ....................... 183
    4.2.3 Multiplicity Variants ......................... 184
  4.3 Leading-Logarithmic Analysis ..................... 187
    4.3.1 Review of Additive Observables ............. 188
    4.3.2 Soft Drop Multiplicity ....................... 192
Acknowledgments

These five years of graduate school have made up one of the most fun and formative chapters in my life. My wife Jie and I moved to Cambridge just a month after getting married, bringing along our son Logan. Having gone to college just 25 minutes away from my childhood home in Florida and living with my parents the whole time, the beginning of graduate school marked many firsts for me. First time husband, first time father, first time being responsible for my own household, first time with a long-term salary. I had a lot to learn about life, not to mention physics.

The coursework in my first year of graduate school was strenuous, satisfying, and rewarding. I learned a lot about quantum field theory and general relativity, thanks to my professors and classmates. I have fond memories of many late nights, working on problem sets with my fellow first-year graduate students. I would like to thank Liujun Zou in particular for keeping that experience fun. I should also thank my undergraduate advisor, Costas Efthimiou, for preparing me for that year.

I would like to thank Matthew Strassler for advising me throughout my second year, while my thesis advisor was busy directing five other students. I learned a lot about how to conduct research from Matt, as well as the painful process of bringing a project all the way through to completion. I also loved working on that first project
with Marat Freytsis and Jakub Scholtz, who taught me a lot about particle physics.

Next I would like to thank my advisor Matthew Schwartz for directing my research efforts throughout graduate school, and for guiding me to become a mature and independent thinker. Matt also taught me a lot about jets, and that the most important part of a presentation is its introduction. (My apologies in advance for the introduction to this thesis that you will find below.)

In my third year, it was fun working with Andrew Larkoski, who taught me about Soft Collinear Effective Theory, and with Kai Yan, who blew me away with the proficiency with which she used it.

In the last couple years, I benefited from working with Jesse Thaler and his undergraduate student Kevin Zhou. I also had a blast collaborating with Ilya Feige and Anders Andreassen on my final project. With them, I learned a lot about machine learning and blossomed as a more independent researcher. I enjoyed working with Hofie Hannesdottir in my last year as well. We spent a lot of time banging our heads against the walls of Matt’s office.

In addition to research, I have enjoyed acting as Teaching Fellow for many courses as a graduate student. I learned a lot about teaching from Jacob Barandes. The most memorable course was quantum field theory, and I am grateful to those students for asking lots of interesting questions.

I also enjoyed sharing an office with quite a few students over the years: David Farhi, Achilleas Porfyriadis, Abhishek Pathak, Monica Pate, Bauer Mukhametzhanov,
and Bruno Balthazar. I am confident that Jefferson 567 has been the quietest office in the department.

My sanity would have completely evaporated sometime in the second year had it not been for religiously lifting weights, three times a week, first with Jae Lee and later with Espen Bernton. I also very much enjoyed unwinding during many camping trips with Espen, Logan, and Michael Rowan. Conversations with my siblings Michael, Jeffrey, and Nicole, and my friend Paul Steidle helped in this way too. I would also like to thank my parents, Robert and Donna, who have been a major source of advice over the years. My mother, and my mother-in-law Qi Zheng, have offered much needed manpower since the birth of my daughter as well.

Through everything, my wife Jie has been by my side, accompanying me through life these last five years. I have benefited a lot from her sense of humor, as well as her lack of sympathy when I complained about something difficult. Two years ago, we welcomed our daughter Jean into the world, who has been smiling and giggling ever since. I have thoroughly enjoyed spending time with Jean, as well as my son and hiking partner Logan, when taking breaks from physics.
Citations to previously published work

Parts of this dissertation cover research reported in the following articles:


Introduction
1.1 Many Useful Theories

The scientific method has guided mankind’s accumulation of trustworthy knowledge for centuries. The method places a large emphasis on falsifiable prediction as one of its cornerstones, and this is the primary reason for its success. Indeed, it is difficult to argue against a theory that forecasts numerous one-in-a-million occurrences correctly, and this is an under exaggeration of what the field of particle physics has done. Using the Standard Model of Particle Physics, the magnetic dipole moment of the electron, for example, has been calculated to 13 decimal places — all of which are in agreement with the experimentally measured value.

Does this mean the Standard Model is the only true theory of nature? After all, the quarks, electrons, gluons, and photons that it describes make up atoms, cells, and mountains. The fields of non-relativistic quantum mechanics, biology, and geology that are used to describe these larger structures were around long before the Standard Model; thus they could not take these fundamental constituents into account. Does that mean these older theories are incorrect?

In a naive sense, yes, those theories are wrong. If one looks too closely at a mountain, say by aiming a particle accelerator at it, one will notice that the rocks are indeed full of quarks and gluons. And if one measures the energy levels of an atom too precisely, one will find them to be in imperfect agreement with a precision quantum mechanics calculation.
But in a much more useful sense, no, the theories of quantum mechanics, biology, and geology are not wrong. Instead, they have a limited scope of applicability. That is, there are certain momentum, energy, length, and time scales on which those theories are designed to make predictions. To put it simply, geology does not correctly predict the outcome of aiming a particle accelerator at a mountain, because that is not really a “mountain question”.

In fact, all scientific theories, including the Standard Model, have a limited scope of practical applicability. There are just too many quarks and electrons involved in the process of cellular reproduction, for example, to take them all into account. It is much easier, and certainly accurate enough, to instead use concepts from biology to describe this process.

Moreover, even if one could simulate all the quarks and electrons involved, thereby taking a purely reductionist view of cellular reproduction, could one really claim to understand it? One of the major goals of science is to chase our human curiosity and desire for understanding. Tracing back a certain effect to a handful of contributing causes and deriving a parameterization of the quantitative importance of those causes allows one to really grasp the essence of the effect in question.

Thus, the many fields of science, each with its limited scope, are incredibly useful in describing natural phenomena. This is not only because each theory provides the most computationally efficient method of calculating the effects it describes, but also because it gives its practitioners the most compelling explanation of those effects.
1.2 Effective Field Theories in Particle Physics

There is a similar story inside the field of particle physics as well. The Standard Model of Particle Physics is the theory that correctly describes matter and its interactions all the way down to the smallest length scales that have ever been probed at modern colliders. It includes particles called quarks, gluons, leptons, neutrinos, the Z and W bosons, the photon, and the Higgs boson. As of summer 2012, when the Higgs was discovered at the Large Hadron Collider, the existence of every particle in the Standard Model has been verified in experiments. This theory thus represents, to the very best of human knowledge, the stuff that everything on Earth is made of, as well as the laws governing the behavior of that stuff.

But even within the field of particle physics, it is rarely the case that in order to accurately describe a certain effect, one must take into account all the particles of the Standard Model. There are just too many of them and, luckily, many of the particles are too heavy or too weakly coupled to be relevant for a given process occurring at a certain energy scale. In most cases, Effective Field Theories (EFTs) become incredibly useful tools. In an EFT, one includes only those degrees of freedom that are essential in accurately describing a certain phenomenon. There are technical methods, referred to as “integrating out degrees of freedom” or “matching”, that unambiguously determine an EFT’s structure. These techniques ensure that the EFT makes predictions in agreement with predictions the full Standard Model would make, if the more compli-
cated calculations were carried out instead.

Physicists were using EFTs even before the concept was developed in the 1970s. To give a specific example from the 1930s, Enrico Fermi’s description of muon decay used an EFT that included the muon, electron, and two neutrinos. Using modern terminology, Fermi’s theory could be obtained from the Standard Model by integrating out the W boson.

At first glance, it may seem surprising that Fermi’s theory was developed decades before the existence of the W boson was postulated. But upon second thought, this is to be expected. In a broad sense, all scientific theories are Effective Field Theories compatible with the Standard Model, and most were developed long before the Standard Model was conceived. The human endeavor of physics began, of course, at the human scales of feet, pounds, and seconds. Physicists slowly worked their way up, to the length scales of the observable universe, as well as down, to the length scale over which the W boson can propagate. Since the muon lives longer and travels farther than the W boson, of course physicists learned to make predictions about muons before W bosons. And insofar as those predictions were correct, of course they were consistent with the more fundamental theory of the Standard Model.

So, the multitude of Effective Field Theories that exist in science should be no surprise. The fact that EFTs work made scientific progress possible in the first place. This fact allows us to understand nature little by little, instead of all at once.

The obvious question that remains is whether the Standard Model itself is yet an-
other EFT of a more fundamental theory. We know for sure there are certain things the Standard Model leaves out. First there are the nonzero neutrino masses; we know these exist but do not know the mechanism that produces them. Next there is dark matter, which is only known to interact gravitationally with the particles of the Standard Model. Third there is the gravitational force itself; while string theory may be a consistent quantum theory of gravity, nobody knows which quantum theory of gravity is actually realized in nature.

So we know the Standard Model is incomplete, being a very successful EFT of some more fundamental physics. The Standard Model contains all the physics necessary to describe all experiments carried out on Earth, down to length scales of $10^{-19}$ meters. To test this, we have accelerated elementary particles at modern colliders to energies over $10^3$ GeV. The major problem facing particle physicists today is that we are not sure where we should look for more fundamental physics. We do know that a quantum theory of gravity will be necessary to correctly describe physics at $10^{19}$ GeV, but we will not be building a collider at that energy scale any time soon. And the physics underlying neutrino masses could lie at any scale in between.

So, in addition to technological difficulties in probing much shorter length scales, for the first time in recent history we do not have a guarantee that there is new physics to be found, even if we built the highest energy machine that current technology could support.

So particle physicists today must make choices about how to spend their time.
Some will find it more compelling to continue studying what new physics might exist, at which energy scales, and how to find it in current or near-future experiments. Others will find it more interesting to instead study interesting emergent phenomena already known to be present in the Standard Model. A third set will manage to keep up with cutting-edge progress in both endeavors.

### 1.3 Jets at Particle Colliders

Of the interesting emergent phenomena contained in the Standard Model, none are more abundantly relevant at modern colliders than jets. Jets are relevant for a number of reasons, several of which are listed below.

First, jets are ubiquitous at modern colliders, and they should be fully utilized so that particle physicists can take advantage of all the data being produced. Let us consider the Higgs boson, for example, which was discovered in 2012 through its decay chains that did not involve jets, but instead electrons, muons, and photons. These no-jet channels include only a small fraction of the total Higgs bosons produced. Indeed, the Higgs boson can undergo a variety of decays: about 60% of the time to quarks, 21% to W bosons, 9% to gluons, 6% to tauons, 3% to Z bosons, and 0.2% to photons. If the Higgs decays to quarks or gluons, these evolve into jets before hitting the detector. But that’s not all. A pair of W bosons involves at least one quark in 94% of its decays, so most of the $\text{Higgs} \rightarrow \text{WW}$ decays show up as jets in the detector as well.
A similar story holds for the tauons and Z bosons. Altogether, 98% of Higgs decays involve jets, but only the remaining 2% were used in its discovery. The jet channels were avoided due to their messiness. Jets involve a lot of particles, and it can be difficult to distinguish jets of different origins, so backgrounds to a Higgs → jets signal are very large.

Second, jets are important in the search for new physics. The clean channels that involve electrons, muons, photons, or missing energy have been explored somewhat thoroughly, but hints of new physics at the LHC have not yet been found. Just as for Higgs physics, if jets are understood well enough to use them in new-physics searches, this will result in a large increase in the amount of data at our disposal.

It seems that higher precision will guide the remainder of the LHC’s runs, whether it be in precision measurements of Standard Model quantities or a precision search for physics beyond the Standard Model. In either case, high precision requires as much data as possible, and much of the data available contains jets.

So, this shows that jets are undeniably useful. Furthermore, they are also an interesting phenomenon in their own right, and they provide another example of where EFT can be very useful.

As mentioned above, jets are messy because they contain so many particles. The average number depends on the energy scale, but it can frequently be several dozen. This poses a problem for precision predictions, which rely almost exclusively on perturbation theory. In perturbation theory, a leading order calculation describing the
high-energy collision is quite easy to perform; it might take someone a minute or an hour, depending on their experience level. But the leading order (LO) calculation allots only one particle for each jet. The next-to-leading order (NLO) calculation includes one extra emission, so that one of the jets involved contains two particles. This alone makes the calculation much more difficult. Realistic NLO computations typically require computer programs and special techniques, which may take of-order years to develop and perhaps a day to run. At next-to-next-to-leading order (NNLO) one of the jets may contain up to 3 particles. This level of precision, however, requires cutting-edge techniques and a given process may take years to become available. For most of the relevant processes, NNLO is state of the art, but it comes nowhere close to describing jet substructure accurately.

For this, a technique called resummation is necessary. Resummation involves taking advantage of patterns in the perturbation series that allow one to sum up an infinite sub-series of terms. These infinite sub-series can describe any number of emissions in a jet, instead of being limited to just a few.

The most systematic way to perform resummation for jet physics is using an EFT, namely Soft-Collinear Effective Theory (SCET). In this theory, the relevant degrees of freedom are quarks collinear to the jet directions, gluons collinear to the jet directions, or soft gluons propagating in any direction. This simplifies calculations with the help of factorization theorems. A factorization theorem might assert that a given observable’s distribution can be calculated as a product of several factors, where each
factor is only sensitive to a single type of mode in the effective theory. This simplifies
the calculation of each factor, and makes resummation, which is well understood in
single-scale problems, doable as well.

The study of jets as an emergent phenomenon using SCET is also a satisfying en-
deavor in its own right. For all the above reasons, jet physics is an interesting field
with a lot to offer.

1.4 Overview of Dissertation

This dissertation contains the majority of my work as a graduate student in chrono-
logical order.

In the second chapter, several precision collider observables are proposed with re-
duced theoretical uncertainties. This is done by taking advantage of symmetries in
the production of electroweak bosons, and by carefully restricting the contribution of
jets. In particular, ratios of diboson production rates are formed, and many uncertain-
ties largely cancel between numerator and denominator, provided the presence of jets
at NLO controlled.

In the third chapter, the problem of precision in jet substructure is attacked head
on. There, a factorization theorem is developed that allows a high precision SCET
calculation of the jet mass. This factorization theorem is made possible by a special
jet grooming technique that removes soft contamination from jets. Such soft con-
tamination has limited the level of resummation that could be performed in the past. The jet groomer, called soft drop, removes soft radiation from the jet with a simple prescription that is analytically tractable. This chapter contains a calculation of the groomed jet mass, resummed to next-to-next-to-leading logarithmic (NNLL) accuracy.

In the fourth chapter, soft drop is used again, here to understand the origin of different jets. A new observable is introduced, called the soft drop multiplicity, which effectively counts the number of hard emissions during the jets evolution. While counting observables are generally difficult to compute analytically, this one admits next-to-leading logarithmic resummation and good analytic understanding. This observable also turns out to be useful in discriminating different types of jets — those originating from quarks and those initiated by gluons. This classification problem is fundamental in QCD. While not approaching the performance of non-analytic discriminants like charged particle multiplicity or neural network methods, soft drop multiplicity does provide insight into the nature of the problem.

In the fifth and final chapter, I present my most recent work, in which I depart from analytical methods, instead using modern machine learning to gain insight about jets. While machine learning methods are relatively new in the field of jet physics, they traditionally provide powerful computational tools, but without much new physical insight into the problems the machine can learn to solve. In this final work, this paradigm is turned around through the use of a Recurrent Neural Network with architecture that is tailor made for jet evolution. This architecture allows for interpreta-
tion through pictures of jet trees and images, allowing its users to gain insights about high-performance discrimination and jet generation.
2

Precision Diboson Observables for the LHC
2.1 Introduction

With no signs as yet of physics beyond the Standard Model (SM), it is essential that measurements at the Large Hadron Collider (LHC) become increasingly precise in the coming years, allowing tests of new SM effects and leading to greater sensitivity to subtle non-SM phenomena. In many cases the limiting factor is a lack of confidence in theoretical calculations, so it is particularly important to find more examples of measurable quantities that are widely agreed to have small theoretical uncertainties.

In this paper we consider production of pairs of electroweak (EW) bosons, collectively referred to as “diboson processes” or $pp \rightarrow V_1 V_2$, where $V_i = \gamma, W^\pm, Z$, which have by now been an object of study for almost four decades [42, 92, 93, 102, 105, 144, 159, 183, 185, 202, 267, 270, 280–283, 297]. These processes have been measured individually by the ATLAS and CMS collaborations [1, 3, 8, 9, 12, 14, 120, 121, 138, 229–231]. Our goal here is to consider combinations of these measurements.

In the SM the EW bosons originate from a triplet and singlet of $SU(2) \times U(1)$, becoming massive and mixing after EW symmetry breaking. But at the high energies accessible to the LHC, the symmetry breaking effects are moderated, and one might imagine the underlying $SU(2) \times U(1)$ structure might more directly relate diboson processes to one another. It turns out that although this naive expectation is not automatically satisfied, there are nevertheless some elegant and interesting relations.

In this paper we identify numerous independent ratios of diboson measurements
that are special at tree level and that offer moderate to excellent potential for both high-precision predictions and high-precision measurements. These ratios, in contrast to the differential cross sections themselves, are flat or slowly-varying as functions of $p_T$ (and other kinematic variables), making them stable against certain experimental problems. Moreover, we expect that many of them receive controllable QCD corrections, especially at high $p_T$. Electroweak corrections are expected to be important at the 10–20% level, and may be visible in these ratios, without clutter from large QCD uncertainties. Since the uncertainties on these EW corrections will be small after ongoing calculations are completed, the ratios potentially also offer sensitivity to high-energy beyond-the-Standard-Model (BSM) phenomena. These would include BSM corrections to triple-gauge-boson vertices and broad diboson resonances, though we do not investigate this issue carefully here.

To illustrate these features, we will perform a detailed study of three related ratios, each of which has a different pattern of uncertainties, though only two of the central values are independent. We will show that their special properties survive to higher order, though with an interesting array of subtleties. Specifically we will consider $d\sigma/d\tilde{m}_T$ for $\gamma\gamma$, $Z\gamma$ and $ZZ$ at next-to-leading order (NLO), where $\tilde{m}_T$ is the average transverse mass of the two vector bosons:

$$\tilde{m}_T = \frac{1}{2} \left( \sqrt{p_{T,1}^2 + m_1^2} + \sqrt{p_{T,2}^2 + m_2^2} \right).$$

We will discuss issues arising at NNLO, and include the $gg$-initiated loop contribution
explicitly. We will give evidence that a number of uncertainties are reduced by taking
the various ratios of these three processes, and also argue that experimental technical-
ities do not interfere with the measurements. The effect of higher-order corrections on
our other observables will be studied elsewhere.

The use of ratios of measurements to reduce theoretical and experimental errors
has a long history, with perhaps the most famous and successful in particle physics
involving the measurements of $R_{\text{had}} = \sigma(e^+e^- \rightarrow \text{hadrons})/\sigma(e^+e^- \rightarrow \mu^+\mu^-)$ in the
early 1970s. In the study of hadronic decay processes, ratios have long been used to
reduce systematic uncertainties from higher-order and non-perturbative corrections
(see ref. [119] and references therein). These methods have seen continuing use at the
LHC, and similar approaches have been extended to the study of Higgs decays in or-
der to better constrain its properties [44, 160, 203].

Ratios of production cross sections at hadron colliders have seen more limited use
due to the more complex initial state. At the LHC in particular, the use of $d\sigma(\gamma + nj)$
to calibrate the process $d\sigma(Z+nj)$, an irreducible background for many BSM searches,
has been investigated at leading order (LO) for $n = 1$ [38] and NLO for $n = 2$ and
3 [77, 78], and implemented in an analysis by the CMS collaboration [124]. Similar
studies have been carried out for ratios of $Z$ and $W^\pm$ processes [261]. Moreover, data
comparing $Z$ to $\gamma$ production has recently been shown to be in good agreement with
theoretical predictions [228], and ratios of single-boson production cross sections have
been measured [4, 11], primarily to aid with fits for parton distribution functions.
Searches for new colored states in ratios of multijet processes have been proposed in ref. [58], while the gradual ramp-up of beam energies at the LHC has also motivated looking at total cross sections of individual processes across a range of energies [263]. More recently it has been argued that a very precise measurement of the top quark Yukawa can be obtained from the ratio of $t\bar{t}h$ to $t\bar{t}Z$ production [262].

2.2 Executive summary

The restoration of $SU(2) \times U(1)$ well above $m_Z$, along with some happy accidents, leads to some interesting relations among the various diboson partonic differential cross sections. These are obscured once the partonic processes are convolved with parton distribution functions (PDFs), and are affected by experimental realities that impact photons, $W$s and $Z$s differently. Nevertheless, at LO we find numerous ratios of differential cross sections for LHC diboson production that have the potential to be interesting observables.

In section 2.3 below, we investigate possible diboson variables at LO. We show that diboson processes naturally divide up into three classes:

\begin{equation}
(1) \gamma\gamma, Z\gamma, ZZ, \quad (2) W^+\gamma, W^\pm Z, \quad (3) W^+W^-.
\end{equation}

(We do not consider same-sign $W^\pm W^\pm$ processes here since extra jets must accompany them.) Each of the first two classes is self-contained, and observables can be built by taking ratios of various differential cross sections. The $W^+W^-$ process can
be related to linear combinations of processes in the first two classes, but is more complicated theoretically.

Our observables involve differential cross sections for $V_1V_2$ production binned in various kinematic variables, which we loosely denote $\sigma(V_1V_2)$ here for brevity. We are interested in symmetric and antisymmetric combinations $\sigma_S$ and $\sigma_A$; here the asymmetry is taken with respect to reversing the relative pseudorapidity $\Delta\eta \equiv \eta_1 - \eta_2$ of the two bosons, signed relative to their longitudinal boost direction. (That is, events are weighted by sign($y_{12}\Delta\eta$), where $y_{12} \approx \frac{1}{2}(\eta_1 + \eta_2)$ is the diboson rapidity. See section 2.3.4 for more details.) We propose that the following ratios are of interest:

1. $R_{1a} = \frac{\sigma_S(Z\gamma)}{\sigma_S(\gamma\gamma)}$, $R_{1b} = \frac{\sigma_S(ZZ)}{\sigma_S(\gamma\gamma)}$, $R_{1c} = \frac{\sigma_S(ZZ)}{\sigma_S(Z\gamma)}$,

2. $C_{2a} = \frac{\sigma_S(W^+\gamma)}{\sigma_S(W^-\gamma)}$, $C_{2b} = \frac{\sigma_S(W^+Z)}{\sigma_S(W^-Z)}$, $D_{2a} = \frac{\sigma_A(W^+\gamma)}{\sigma_A(W^-\gamma)}$, $D_{2b} = \frac{\sigma_A(W^+Z)}{\sigma_A(W^-Z)}$,

3. $R_{1c} = R_{1b}/R_{1a}$, $R_{1b}^+ = R_{1a}/R_{1b}$, $A_{2a} = C_{2b}/C_{2a}$, $A_{2b} = D_{2b}/D_{2a}$ — the pattern of theoretical and statistical uncertainties is different for each ratio.

where $V^0$ denotes $Z$ or $\gamma$, and $\sigma_A(WV^0)$ is some linear combination of $\sigma_A(W^+V^0)$ and $\sigma_A(W^-V^0)$. See section 2.3.5 for a more precise discussion of $R_3$ and $A_3$.

In figures 2.3–2.6 of section 2.3.5, these ratios, calculated at LO and binned in $s$, are shown. All of the ratios are slowly varying, and each has its own special features.

---

1. Although the central values of these observables are not all independent — for instance $R_{1c} = R_{1b}/R_{1a}$, $R_{1b}^+ = C_{2b}/C_{2a}$, $A_{2a}^+ = D_{2b}/D_{2a}$ — the pattern of theoretical and statistical uncertainties is different for each ratio.
Observables $R_{1a}$, $R_{2}^{\pm}$, and $A_{2}^{\pm}$ are, to first approximation, independent of the PDFs (and hence have very small PDF uncertainties). At LO they depend only on ratios of SM couplings and charges, from which we learn $R_{1a}$ is nearly constant, $R_{2}^{+} \approx R_{2}^{-}$, and $A_{2}^{\pm} \approx -1$. By contrast, observables $R_{1b}, R_{1c}, C_{2a}, C_{2b}, D_{2a}, D_{2b}$ are dominated by the difference between up and down PDFs; all SM couplings cancel in the $C_{2}$ and $D_{2}$ ratios. Observables $R_{3}$ and $A_{3}$ are more complex.

These observables are simplest for $\sqrt{s} \gg 2m_{Z}$ or $m_{T} \gg m_{Z}$, where the difference between the massless $\gamma$ and the massive $W, Z$ is of diminished importance. But as discussed in section 2.3.6, the low production rates for diboson processes at these high scales, and the low branching fraction for $Z \rightarrow$ leptons, gives our observables relatively large statistical uncertainties, potentially negating the value of their low theoretical uncertainties. (In this paper we will only consider leptonic decays of $W$s and $Z$s, though we briefly discuss other options in section 2.6.2.) At 300 fb$^{-1}$, the $R_{1a}, C_{2a}$ and $R_{3}$ observables can be measured in multiple bins with 5% statistical uncertainties. This is comparable to the theoretical uncertainties that we will claim below. The variables $R_{2}^{\pm}$ and $D_{2a}$ can only be measured in a single bin, making them only marginally useful. At 3000 fb$^{-1}$, it appears all the variables are potentially useful excepting only $D_{2b}$ and $A_{2}^{-}$, and with $A_{2}^{+}$ marginal.

In section 2.4, we study the simplest of these observables, the $R_{1}$ ratios, beyond LO. As described in section 2.4.1, we choose our cuts and our observable carefully to avoid strong jet vetoes, problematic kinematic regions with very large $K$ factors, etc.;
see table 2.3 and table 2.5 below. We also include $gg$ production, formally NNLO but numerically important. To fix its normalization, we use the fact that the dominant correction to $gg \to \gamma\gamma$ at the next order is known \cite{80}. We also use this to normalize the other $gg \to V_1^0V_2^0$ processes.\footnote{As this paper was nearing completion, a calculation for $gg \to ZZ$ analogous to ref. \cite{80} appeared in ref. \cite{106}. Our normalization estimate appears to agree with their results.}

In section 2.4.2, we show that many NLO QCD corrections do cancel in these ratios, except for the region where a final-state jet is collinear with a vector boson. There the photon has a collinear singularity which must be regulated with, e.g., a fragmentation function, while the $Z$ singularity is regulated by its mass. Although the ratios shift significantly in this region, we argue in section 2.5.1 that use of a “staircase” isolation method, as in ref. \cite{88, 212}, leaves small theoretical uncertainties. We also show in section 2.4.3 that $gg \to V_1^0V_2^0$ causes shifts in the ratios as large as 5–20\% at low $\bar{m}_T$, due in part to an interesting accidental cancellation in $gg \to Z\gamma$, though these effects are reduced at high $\bar{m}_T$. Moreover, we argue that the uncertainties on these shifts are small. We also discuss other known NNLO effects on our ratios. Finally, we find in section 2.4.4 that certain other QCD theoretical uncertainties — PDF uncertainties and scale uncertainties in particular — do largely cancel, especially for $R_{1a}$.

These statements are summarized in figure 2.1. To explain this figure, let us focus first on the top plot, which shows results for $R_{1a}$, the ratio of $Z\gamma$ to $\gamma\gamma$ differential cross sections with respect to $\bar{m}_T$, obtained for the 13 TeV LHC. The upper portion
Figure 2.1: (Top) $R_{1a} = \sigma_S(Z\gamma)/\sigma_S(\gamma\gamma)$. (Left) $R_{1b} = \sigma_S(ZZ)/\sigma_S(\gamma\gamma)$. (Right) $R_{1c} = \sigma_S(ZZ)/\sigma_S(Z\gamma)$. The solid symbols represent our NLO (+ NNLO $gg$) theoretical prediction. Their error bars indicate the expected statistical uncertainties after 300 (3000) fb$^{-1}$ for $R_{1a}$ ($R_{1b}$ and $R_{1c}$). The shaded band around these points represents our estimate of QCD theory uncertainties; see text for important details. The corresponding LO theory prediction is given in open symbols. (By chance, higher-order corrections to $R_{1c}$ nearly cancel.) The bottom plot for each ratio shows the expected fractional correction (relative to unity) from additional non-QCD corrections: an orange solid line for the effect of $Z \to \ell\ell$ decays on the experimental measurement, a blue dashed line for an estimate of the effect of electroweak Sudakov logarithms, with a band indicating its uncertainty, and a horizontal band for the uncertainty from the undetermined choice of $\alpha_{\text{QED}}$. 
of the plot shows the ratio $R_{1a}$ as would be measured in 6 bins of 5–6% statistical uncertainty; the last bin includes events with $\bar{m}_T$ extending up to the kinematic limit. The open circles indicate a LO prediction, while the closed circles are our result including NLO and $gg$-initiated production. The dominant corrections are driven by the gluon PDF, and decrease with $\bar{m}_T$. The error bars on the closed circles indicate the expected statistical errors at 300 fb$^{-1}$. The shaded band indicates the theoretical uncertainties mentioned in the previous paragraphs, itemized in table 2.6 of section 2.6 and with all uncertainties combined linearly, except for PDF extraction uncertainties which are combined in quadrature with the others. This combination gives a conservative estimate of known uncertainties.

We emphasize that we have not proven it impossible for additional unknown sources at NNLO to shift the ratios’ central values by larger amounts than our uncertainty estimates. Although we believe we identified all obvious effects that do not cancel in ratios, and have either included them or estimated our uncertainties from not including them, we cannot demonstrate this directly. Only the complete NNLO calculations, for which code is not yet public, will confirm that there are no additional subtleties.

The lower portion of the plot shows estimates of three sources of additional corrections and their uncertainties, expressed as a relative shift of the ratio; (i.e. 1.05 indicates an upward shift of 5% on the ratio.) First, as discussed in section 2.4.5.1, leading-log EW corrections only partially cancel in the $R_1$ ratios. At high $\bar{m}_T$ Sudakov logarithmic effects will dominate and can be roughly estimated using the soft-
collinear approximation, as studied in ref. [62]. The effect on $R_{1a}$ arises as a difference between the $Z$ and $\gamma$ jet functions, and is of order 5–10% at high $\bar{m}_T$, though this is probably an overestimate. We show this estimate by plotting the effect on our ratios of the calculation of ref. [62] as a blue dashed line, along with an estimate of its uncertainty band as a shaded blue region. At low $\bar{m}_T$ a finite correction, still relatively small, may make the true EW shift of $R_{1a}$ somewhat larger than indicated by our blue band — see [87, 158], although their cuts are significantly different from ours. Nevertheless, and more importantly, our uncertainty band is conservative. The band correctly shows the dominant uncertainty at high $\bar{m}_T$, from matching the resummed and fixed-order calculations. At small $\bar{m}_T$ the leading uncertainty, from scale variation of the EW couplings, is smaller than the band.

Second, the tan horizontal shaded bar represents an unresolved disagreement in the community, discussed in section 2.4.5.2, regarding the choice of scale $\mu$ for evaluating $\alpha_{\text{QED}}$ when an on-shell photon is emitted in a hadronic setting. The difference between using $\mu = 0$ and $\mu = m_Z$ — for each observable, an overall shift of all the bins by a nearly equal amount — is indicated by this bar. This issue is temporary; the uncertainty will be eliminated once the controversy is settled.

Third, we have chosen to show our results in the upper portion of the figure without including effects from $Z$ decays to leptons. That is, in the figure we applied cuts on the vector bosons but ignored the finite $Z$ width and the kinematic and isolation cuts that must be imposed on the leptons. As we study in section 2.5.2, these effects,
shown as an orange solid line in the lower portion of the figure, do materially change the ratios at the $\sim 5$–15% level, but with very low uncertainty.

In the other two plots of figure 2.1, we show similar results for $R_{1b}$ and $R_{1c}$, but at 3000 fb$^{-1}$. The increased integrated luminosity is required in order to obtain small statistical errors, because of the small branching fraction of $ZZ$ to four leptons. Both QCD and EW corrections to $R_{1b}$ are larger because the differences between $Z$ and $\gamma$ contribute twice.

We see from figure 2.1 that the variables $R_{1a}$, $R_{1b}$ and $R_{1c}$ are nearly flat in $\bar{m}_T$, are potentially predictable at better than 5%, and are measurable in several bins (using only leptonic $Z$ decays) at the $\sim 5$–6% level with 300, 3000 and 3000 fb$^{-1}$ respectively. Corrections to the LO prediction are moderate at low $\bar{m}_T$ and decrease with $\bar{m}_T$. (In $R_{1c}$ the prediction at higher-order is nearly the same as at LO, due to an accidental cancellation between the $gg$ contribution and other corrections.) Moreover, at 3000 fb$^{-1}$ the $R_{1a}$ ratio can be measured using tens of bins (the precise number depending on $\bar{m}_T$ resolution) with the highest bin starting above 600 GeV, nearly double what is possible at 300 fb$^{-1}$.

At this level of precision, these ratios are potentially sensitive both to interesting soft-collinear EW corrections and to BSM phenomena. We are optimistic that other variables in our list will prove comparably useful, though this remains to be shown in future work.
2.3 The story at leading order

We begin with a study of diboson processes at tree level, which were first computed at this order almost four decades ago [92, 93, 270]. In the form originally presented, the underlying broken gauge and custodial symmetries were not manifest. Making these more explicit, we identify ratios of particular interest. As we will see, each ratio has its own unique features, strengths and weaknesses, even at leading order. We will study these features first at the partonic level, where the $SU(2) \times U(1)$ structure of the rates is most clear. We then use this structure as a guide to construct our ratio observables. Finally we show and explain the behavior of these ratios in proton-proton collisions at 13 TeV. We conclude this section with a short discussion of the statistical uncertainties on these variables at 300 and 3000 fb$^{-1}$ at 13 TeV.

2.3.1 High energy limit

Well above the scale of EW symmetry breaking, we may rewrite the SM EW bosons $W^\pm, Z, \gamma$ as the triplet $w^\pm, w^3$ and singlet $x$ of massless gauge bosons of $SU(2) \times U(1)$, along with the Goldstone scalars $\phi^\pm, \phi^3$. (We use lowercase letters for massless gauge bosons and capital letters for the mass eigenstates.) One basis for the massless diboson states consists, up to normalizations, of $SU(2) \times U(1)$ singlets and triplets:

\begin{equation}
xx_1 \equiv xx : \ |xx\rangle , \tag{2.3.1}
\end{equation}

\begin{equation}
wx_3 \equiv wx : \ |w^+x\rangle , \ |w^3x\rangle , \ |w^-x\rangle , \tag{2.3.2}
\end{equation}

25
Figure 2.2: At leading order, diboson processes proceed from $qar{q}$ initial states. The $t, u$ channels (left) and the $s$ channel (right) contribute only to particular amplitudes under $SU(2) \times U(1)$.

\[ ww_1 : \ |w^+w^-| + |w^-w^+| - |w^3w^3| , \quad (2.3.3) \]

\[ ww_3 : \ |w^+w^3| - |w^3w^+| , \ |w^+w^-| - |w^-w^+| , \ |w^3w^-| - |w^-w^3| . \quad (2.3.4) \]

There are also quintet $ww$ states, such as $W^+W^+$, but they require two final-state jets at LO, whereas we will focus on production with no jets at LO. This means we only deal at LO with three $SU(2)$-singlet $qar{q}$ initial states

\[ |u_R\bar{u}_R\rangle , \ |d_R\bar{d}_R\rangle , \ |u_L\bar{u}_L\rangle - |d_L\bar{d}_L\rangle , \quad (2.3.5) \]

and the triplet of states

\[ \{ |u_L\bar{d}_L\rangle , \ |u_L\bar{u}_L\rangle + |d_L\bar{d}_L\rangle , \ |d_L\bar{u}_L\rangle \} . \quad (2.3.6) \]

Production rates at LO involve $s, t, u$-channel Feynman diagrams; see figure 2.2. The $s$-channel diagram, with an $f^{abc}$ symbol, only contributes for $ww_3$ states. Because of this, the LO production rates for $xx, wx$, and $ww_1$ are proportional, differing only in the coupling constants.

This suggests that symmetries should exist among the observable cross sections of interest $\sigma(pp \rightarrow V_1V_2)$. To determine the implications more precisely, we must take into account the production of scalars (e.g., the $\phi^3$ inside $Z$), the interference
between different channels (e.g., since $W^{-\gamma}$ is a superposition of $wx$ and $ww_3$), and the convolution with PDFs.

Since the quark-scalar couplings are proportional to quark masses, we can neglect scalar production in the $t$- and $u$-channel diagrams, so the scalars contribute only to triplet processes. When final-state scalars do contribute, they do so in the spin-sum of squared helicity-amplitudes, so there are no associated interference effects.

### 2.3.2 Squared amplitudes

The production of dibosons in the limit in which their masses can be neglected can be written in a simple form. We will denote the coupling-stripped LO singlet-, triplet- and scalar amplitudes by

\begin{align}
  a_1 &\propto \mathcal{M}(xx) \propto \mathcal{M}(wx) \propto \mathcal{M}(ww_1), \\
  a_3 &\propto \mathcal{M}(ww_3), \\
  a_\phi &\propto \mathcal{M}(\phi\phi),
\end{align}

in a notation which corresponds to eqs. (2.3.1)–(2.3.4). In these schematic definitions, we leave polarizations implicit since we will always compute spin-averaged cross sections. The three amplitudes in the first line are all proportional, and this continues to hold when one includes NLO QCD corrections but not NLO EW corrections.\(^3\)

3For instance, a virtual $w$ can attach to the final-state lines in $\mathcal{M}(ww_1)$ but not in $\mathcal{M}(xx)$.
quadratic in the $a_i$s. The products of $a_i$s that are relevant for diboson production include

$$|a_1|^2 = \frac{\hat{t}}{\hat{u}} + \frac{\hat{u}}{\hat{t}}.$$  
(2.3.10)

$$a_1 a_3 = \left( \frac{\hat{t} - \hat{u}}{2s} \right) - \frac{1}{4} \left( \frac{\hat{t}}{\hat{u}} - \frac{\hat{u}}{\hat{t}} \right),$$  
(2.3.11)

$$|a_3|^2 = \frac{\hat{t} \hat{u}}{4s^2} - \frac{1}{8} \left( \frac{\hat{t}}{\hat{u}} + \frac{\hat{u}}{\hat{t}} \right),$$  
(2.3.12)

$$|a_\phi|^2 = \frac{\hat{t} \hat{u}}{4s^2}. $$  
(2.3.13)

Here, $(a_1 a_3)$ is shorthand for $\text{Re}(a^*_1 a_3)$. The $a_i$ amplitudes transform simply under

$$\hat{t} \leftrightarrow \hat{u} \text{ exchange:}$$

$$a_1(\hat{t}, \hat{u}) = a_1(\hat{u}, \hat{t}), \quad a_3(\hat{t}, \hat{u}) = -a_3(\hat{u}, \hat{t}), \quad |a_\phi(\hat{t}, \hat{u})| = |a_\phi(\hat{u}, \hat{t})|. $$  
(2.3.14)

These properties of $a_1$ and $a_3$, required by Bose statistics and by the fact that $ww_1$ ($ww_3$) is symmetric (antisymmetric) in the two $w$s, explain why in eqs. (2.3.10)–(2.3.13) only $(a_1 a_3)$ is antisymmetric under $\hat{t} \leftrightarrow \hat{u}$.

The $\hat{t} \leftrightarrow \hat{u}$ symmetry properties of the $a_i$s play an important role in what follows. These are forward-backward symmetries, since swapping $\hat{t} \leftrightarrow \hat{u}$ in a $q\bar{q} \rightarrow V_1V_2$ event reverses the sign of $\eta_1 - \eta_2$, with $\eta$ defined relative to the $q$’s momentum direction. In what follows, we will use $d\hat{\sigma}_S$ ($d\hat{\sigma}_A$) to denote $\hat{t} \leftrightarrow \hat{u}$ symmetrized (antisymmetrized)

---

4These expressions can be extracted from the high-energy limit of the partonic rates in Eqs. (A.1.1)–(A.1.6) below, which were computed in refs. [92, 93, 270].

5Notice that NLO EW corrections break the $\hat{t} \leftrightarrow \hat{u}$ symmetry of $\mathcal{M}(wx)$ since a virtual $w$ can attach to the final-state $w$ line but not to the $x$ line.
partonic differential cross sections. We will discuss symmetric and antisymmetric hadronic cross sections $\sigma_S, \sigma_A$ in section 2.3.4.

One important consequence of eq. (2.3.14) is that $a_3$ vanishes at $\hat{t} = \hat{u}$, that is, at center-of-mass-frame (CM) scattering angle $\theta = \pi/2$. This “radiation zero” has an important impact on the diboson processes.

2.3.3 Partonic cross sections at high energies

Next we write the partonic cross sections for the production of physical dibosons $V_1 V_2$, ignoring mass corrections of order $m_Z^2/p_T^2$. Our formulas are written in terms of the $a_i$s given in eqs. (2.3.10)–(2.3.13), making various relations among the cross sections manifest and motivating the ratio observables mentioned in section 2.2.

The full formulas including $O(m_Z^2/p_T^2)$ terms are given in Appendix A.1. There we define $A_i$s as straightforward generalizations of the $a_i$s including mass corrections. These corrections are subleading in the region of phase space we study in this paper compared to certain QCD corrections, and they introduce no uncertainties. We include them in our numerical results, but have no need to discuss them further. In fact a few useful relations, such as eqs. (2.3.22)–(2.3.23), are unaffected by the boson masses.
2.3.3.1 $\gamma\gamma$, $Z\gamma$, $Z\bar{Z}$

Writing $c_W = \cos \theta_W$ and $s_W = \sin \theta_W$, we have

$$\gamma = c_W x + s_W w^3,$$

(2.3.15)

$$Z = c_W w^3 - s_W x,$$

(2.3.16)

and $Z$ also contains the scalar $\phi^3$. Pairs of photons and $Z$s can be produced in $xx$, $w^3x$, and $w^3w^3$ channels. Since $w^3w^3$ is orthogonal to the $ww_3$ states, the production rates in this sector are all proportional to $|a_1|^2$; see eq. (2.3.7). Inserting the appropriate coupling constants and writing $V^0 = \gamma, Z$, we have

$$\frac{d\sigma}{dt}(q\bar{q} \to V_1^0V_2^0) = \frac{C_{\gamma\gamma}^q}{s^2}|a_1|^2,$$

(2.3.17)

where

$$C_{\gamma\gamma}^q = \frac{1}{2} \frac{\pi^2 s_W^4}{c_W^2} 2Q^4,$$

(2.3.18)

$$C_{Z\gamma}^q = \frac{\pi^2 s_W^2 c_W^2}{N_c} (L^2Q^2 + R^2Q^2),$$

(2.3.19)

$$C_{ZZ}^q = \frac{1}{2} \frac{\pi^2 s_W^4}{c_W^2} (L^4 + R^4).$$

(2.3.20)

Here, a symmetry factor of 1/2 has been included for identical particles, $\alpha_2$ is the $SU(2)$ coupling of the SM, $Q = T_3 + Y$ is the electric charge of quark $q$, and

$$L = T_3 - Y_L t_W^2, \quad R = -Y_R t_W^2,$$

(2.3.21)

with $t_W = s_W/c_W$. The $O(m_Z^2/p_T^2)$ corrections to eq. (2.3.17) are given in Appendix A.1.

Each partonic rate in this sector is forward-backward symmetric, so $d\sigma_A(V_1^0V_2^0) = 0$
(though NLO EW corrections give a non-zero $d\hat{\sigma}_A(Z\gamma)$.)

### 2.3.3.2 $W^{\pm}\gamma$, $W^{\pm}Z$

We begin this section by discussing relations among $W^+V^0$ and $W^-V^0$ rates. Since $W^+V^0$ and $W^-V^0$ production are related by $CP$, which takes $u\bar{d} \to W^+V^0$ into $d\bar{u} \to V^0W^-$, we have (in the notation of section 2.3.2)

$$d\hat{\sigma}_S(u\bar{d} \to W^+V^0) = d\hat{\sigma}_S(d\bar{u} \to W^-V^0), \quad (2.32)$$

$$d\hat{\sigma}_A(u\bar{d} \to W^+V^0) = -d\hat{\sigma}_A(d\bar{u} \to W^-V^0). \quad (2.33)$$

Next we write down the partonic cross sections for producing $W^{\pm}V^0$. These arise from $w^{\pm}w^3$ and $w^\pm x$ and involve both $a_1$ and $a_3$, as seen from eqs. (2.3.2)–(2.3.4) and (2.3.7)–(2.3.9). Scalar production $a_\phi$ also appears in $W^{\pm}Z$. In particular,

$$\frac{d\hat{\sigma}}{dt}(q\bar{q}' \to W^{\pm}\gamma) = \frac{\pi|V_{ud}|^2\alpha_2^2s_W^2}{N_c s^2} \left[ \frac{Y_L^2|a_1|^2 - 2Y_L(a_1a_3) + 4|a_3|^2}{2} \right], \quad (2.34)$$

$$\frac{d\hat{\sigma}}{dt}(q\bar{q}' \to W^{\pm}Z) = \frac{\pi|V_{ud}|^2\alpha_2^2}{N_c s^2} \left[ \frac{s_W^2t^2Y_L^2|a_1|^2 - 2s_W^2Y_L(a_1a_3) + 4c_W^2|a_3|^2 + \frac{1}{2}|a_\phi|^2}{2} \right], \quad (2.35)$$

where $q\bar{q}'$ is $u\bar{d}$ ($d\bar{u}$) for $W^+V^0$ ($W^-V^0$). The $O(m_Z^2/p_T^2)$ terms in these rates are given in Appendix A.1. As seen from eq. (2.3.14), these formulas obey eqs. (2.3.22)–(2.3.23).

Next we compare $W^{\pm}\gamma$ to $W^{\pm}Z$. Notice that the forward-backward antisymmetric terms in these two rates, those proportional to $Y_L(a_1a_3)$, are equal but opposite:

$$d\hat{\sigma}_A(W^{\pm}\gamma) = -d\hat{\sigma}_A(W^{\pm}Z). \quad (2.36)$$
These asymmetries arise from the interference between $w^\pm w^3$ and $w^\pm x$ production, a cross term that carries opposite sign for the photon versus the $Z$; see eqs. (2.3.15)–(2.3.16). Alternatively, completeness requires that in the high energy limit,

$$d\hat{\sigma}(W^\pm \gamma) + d\hat{\sigma}(W^\pm Z) = d\hat{\sigma}(w^\pm x) + d\hat{\sigma}(w^\pm w^3) + d\hat{\sigma}(\phi^\pm \phi^3).$$

(2.3.27)

Since the three terms on the right hand side are respectively proportional to $|a_1|^2$, $|a_3|^2$ and $|a_\phi|^2$, which are forward-backward symmetric, eq. (2.3.26) follows.

The forward-backward symmetric rates in this sector can be read from eqs. (2.3.24) and (2.3.25) by omitting the $(a_1 a_3)$ terms. Because of the smallness of $Y_L^2 = 1/36$ and the relative factor of $(8c_W)^{-1}$ suppressing $|a_\phi|^2$, the $|a_3|^2$ terms naively dominate the cross sections, leading to a ratio $d\hat{\sigma}_S(W^+ \gamma)/d\hat{\sigma}_S(W^+ Z)$ of $t_W^2 \approx 0.29$.

However, there is a small subtlety with this estimate. We noted earlier that $a_3$, antisymmetric under $\hat{t} \leftrightarrow \hat{u}$, has a radiation zero.\(^6\) Nonetheless, the coefficients of $|a_1|^2$ and $|a_\phi|^2$ are small, so this zero is only important very close to $\theta \sim \pi/2$. Moreover, by chance, the ratio of $d\hat{\sigma}_S(W^+ \gamma)$ to $d\hat{\sigma}_S(W^+ Z)$ is 0.19 at $\theta = \pi/2$, protecting the naive estimate of $t_W^2$ from a large correction. We will say more about this in section 2.3.5.

### 2.3.3.3 $W^- W^+$

The partonic amplitude for producing transversely-polarized $W^- W^+$ is a linear combination of $a_1$ and $a_3$ in the high-energy limit. One must also include the contribu-

\(^6\)This radiation zero of $a_3$ combines with $a_1$ to give the famous tree-level $f \bar{f}' \rightarrow W\gamma$ radiation zero [270], at an angle that depends on the electric charge of $f$. 

32
tation $a_\phi$ from scalars $\phi^- \phi^+$, which are produced through an $s$-channel $w^3$ or $x$ in $q_L \bar{q}_L$-initiated processes, or through an $s$-channel $x$ from $q_R \bar{q}_R$.

In the high energy limit, the partonic cross sections are

$$\frac{d\hat{\sigma}}{dt}(q\bar{q} \rightarrow W^- W^+) = \frac{\pi \alpha_s^2}{N_c s^2} \left\{ \frac{1}{16} |a_1|^2 + \frac{1}{2} (a_1 a_3) + 2 |a_3|^2 \right.$$ \nonumber 

$$+ \left[ (t_W^2 Y_R)^2 + (t_W^2 Y_L + T_3)^2 \right] |a_\phi|^2 \right\}, \quad (2.3.28)$$

where the upper (lower) sign holds for $u$-type ($d$-type) quarks. Here $T_3, Y_L, Y_R$ are the quantum numbers of quark $q$. Note that the forward-backward symmetric rates for transversely polarized $W^- W^+$ are the same in $u\bar{u}$ and $d\bar{d}$ channels, while the forward-backward antisymmetric rates are equal and opposite; that is,

$$d\hat{\sigma}_S(u\bar{u} \rightarrow W_T^- W_T^+) = d\hat{\sigma}_S(d\bar{d} \rightarrow W_T^- W_T^+), \quad (2.3.29)$$

$$d\hat{\sigma}_A(u\bar{u} \rightarrow W_T^- W_T^+) = -d\hat{\sigma}_A(d\bar{d} \rightarrow W_T^- W_T^+). \quad (2.3.30)$$

These relations are a consequence of $G$-parity (charge conjugation $C$ followed by a rotation by $\pi$ around the second isospin axis) which takes $u\bar{u} \rightarrow w^- w^+$ into $d\bar{d} \rightarrow w^+ w^-$. Indeed, high energy production of $W_T^- W_T^+$ (which in our notation is equivalent to $w^- w^+$) proceeds at LO only through $SU(2)$ interactions, which respect $G$-parity. Alternatively one can derive eqs. (2.3.29) and (2.3.30) using Clebsch-Gordan coefficients:

$$\mathcal{M}(u\bar{u} \rightarrow w^- w^+) = \frac{1}{2} \mathcal{M}(q\bar{q}_3 \rightarrow ww_3) + \frac{1}{\sqrt{6}} \mathcal{M}(q\bar{q}_1 \rightarrow ww_1), \quad (2.3.31)$$

$$\mathcal{M}(d\bar{d} \rightarrow w^- w^+) = \frac{1}{2} \mathcal{M}(q\bar{q}_3 \rightarrow ww_3) - \frac{1}{\sqrt{6}} \mathcal{M}(q\bar{q}_1 \rightarrow ww_1). \quad (2.3.32)$$
Squaring these equations and referring to relations eq. (2.3.14), one finds that eq. (2.3.29) must hold, with \( d\tilde{\sigma}_S \) given by a linear combination of \( |a_1|^2 \) and \( |a_3|^2 \). And since the cross terms have opposite signs, eq. (2.3.30) follows, with \( d\tilde{\sigma}_A \) proportional to \( (a_1 a_3) \).

On the other hand, note that the \( Y_L T_3 \) terms in \( d\tilde{\sigma}(u\bar{u} \rightarrow \phi^- \phi^+) \) and \( d\tilde{\sigma}(d\bar{d} \rightarrow \phi^- \phi^+) \) are not equal even though they are forward-backward symmetric. These terms arise from an \( s \)-channel \( x \) boson, which interacts with the initial-state quarks with couplings that violate \( G \)-parity. However, these terms are numerically small.

Since \( d\tilde{\sigma}_A(W^- W^+) \propto (a_1 a_3) \), the partonic asymmetry of \( W^- W^+ \) is proportional to\(^7\) that of \( W^- \gamma \) and \( W^- Z \). Meanwhile the radiation zero of \( a_3 \) is quite important for \( d\tilde{\sigma}_S(W^- W^+) \). Later we will see that \( |a_1|^2 \) actually dominates the \( W^- W^+ \) cross section, though not overwhelmingly. This motivates comparing \( d\tilde{\sigma}_S(W^+ W^-) \) to \( d\tilde{\sigma}_S(V_1^0 V_2^0) \propto |a_1|^2 \), or perhaps to a linear combination of \( d\tilde{\sigma}_S(V_1^0 V_2^0) \) and \( d\tilde{\sigma}_S(W V^0) \).

### 2.3.4 Convolution with PDFs

Having discussed the partonic cross sections in detail, we now turn to the observable hadronic cross sections

\[
\begin{align*}
\sigma(pp \rightarrow V_1 V_2) &= \sum_{q,q'} dx_1 dx_2 f_q(x_1) f_{q'}(x_2) d\tilde{\sigma}(qq' \rightarrow V_1 V_2) \\
&= \sum_{q,q'} \frac{d\hat{s}}{s} dy f_q(x_1) f_{q'}(x_2) d\tilde{\sigma}(qq' \rightarrow V_1 V_2). \quad (2.3.33)
\end{align*}
\]

\(^7\)But note \( d\tilde{\sigma}_A(W^- W^+) \) arises as interference between \( \mathcal{M}(w w_3) \) and \( \mathcal{M}(w w_1) \), while \( d\tilde{\sigma}_A(W^+ V^0) \) is an interference between \( \mathcal{M}(w w_3) \) and \( \mathcal{M}(w x) \). Since NLO EW corrections break the LO relation \( \mathcal{M}(w w_1) \propto \mathcal{M}(w x) \), they also violate \( d\tilde{\sigma}_A(W^- W^+) \propto d\tilde{\sigma}_A(W^+ V^0) \).
Here $f_i(x)$ is the PDF of parton $i$, $\hat{s} = x_1 x_2 s$ is the CM energy, and $y = \frac{1}{2} \log(x_1/x_2)$ is the rapidity of the partonic collision.

To fully specify an event, kinematic variables describing the final state must be chosen. Since our purpose is to study ratios of different diboson processes, we want variables that keep the different processes on equal footing to the extent possible. One useful variable is $m_{VV}$, the invariant mass of the two bosons; this equals $\sqrt{\hat{s}}$ at LO. Considerations at LO might also suggest the use of the transverse momentum $p_T$ of either boson. However, the threshold value of $\hat{s}$ required to produce the $V_1V_2$ pair with a given $p_T$ differs among the processes:

$$\hat{s}_{\text{thresh}} = \left( \sqrt{p_T^2 + m_1^2} + \sqrt{p_T^2 + m_2^2} \right)^2 = 4\bar{m}_T^2,$$ (2.3.34)

where $\bar{m}_T$ is the average transverse mass of the two final-state bosons. Since our ratios are simpler if partonic kinematics span the same range in numerator and denominator, the above relation suggests that $\bar{m}_T$ is a more useful kinematic variable than $p_T$.

The partonic cross sections $d\hat{\sigma}/dt$ given in section 2.3.3 can be rewritten in terms of $\bar{m}_T$ as

$$\frac{d\hat{\sigma}}{d\bar{m}_T}(qq' \to V_1V_2) = \left| \frac{dt}{d\bar{m}_T} \right| \frac{d\hat{\sigma}}{dt}(qq' \to V_1V_2),$$ (2.3.35)
where, if $m_1 = m_2$ or if both $m_1$ and $m_2$ are negligible,\textsuperscript{8}

$$\left| \frac{d\hat{t}}{d\hat{m}_T} \right| = 2\hat{m}_T \left( 1 - \frac{4\hat{m}_T^2}{s} \right)^{-1/2}.$$  \hspace{1cm} (2.3.36)

The corresponding observable cross section takes the form

$$\sigma(pp \to V_1V_2) = \sum_{q,q'} \int \frac{d\hat{s}}{s} \int d\hat{m}_T \frac{d\hat{\sigma}}{d\hat{m}_T}(qq' \to V_1V_2) \int dy f_q(x_1) f_{q'}(x_2),$$  \hspace{1cm} (2.3.37)

where the domain of integration depends on the observable being computed and the kinematic cuts imposed.

The observables we propose in this paper involve the quantities $\sigma_S$ and $\sigma_A$ which we now define. We have already introduced $d\hat{\sigma}_S (d\hat{\sigma}_A)$ as the $\hat{t} \leftrightarrow \hat{u}$ symmetric (antisymmetric) part of the differential partonic cross section. That is, $d\hat{\sigma}_A(q\bar{q} \to V_1V_2)$ weights events by $\text{sign}(\eta_1 - \eta_2)$, while $d\hat{\sigma}_S$ weights events symmetrically with $+1$. At $pp$ colliders, the $q$ direction is unobservable but is typically aligned with the longitudinal boost $y_{12}$ of the diboson system, which at LO is the same as the boost $y$ of the $q\bar{q}$ center-of-mass frame. We may thus define $\sigma_A$ at LO by assigning to events the weight $\text{sign}[y(\eta_1 - \eta_2)]$, as in

$$\sigma_X^{\text{LO}}(pp \to V_1V_2) = \sum_{q_i,q_j} \int \frac{d\hat{s}}{s} \int d\hat{m}_T \frac{d\hat{\sigma}_X^{\text{LO}}}{d\hat{m}_T}(q_i\bar{q}_j \to V_1V_2) \mathcal{L}_{q_i\bar{q}_j}^X,$$  \hspace{1cm} (2.3.38)

where $X = S, A$ and we have introduced

$$\mathcal{L}_{q_i\bar{q}_j}^{X(S,A)} = \int dy \{1, \text{sign}(y)\} 2f_{q_i}(x_1)f_{\bar{q}_j}(x_2)$$  \hspace{1cm} (2.3.39)

as symmetric and antisymmetric parton luminosities. The limits of integration on $y$

\textsuperscript{8}The Jacobian is considerably more complicated when $m_1 \neq m_2$. 

36
depend on $s$ and $\hat{m}_T$ once cuts are imposed on the pseudorapidity of the bosons.

Triply-differential cross sections would show the relations among the diboson processes most directly, since the PDFs would be evaluated in small $x_1, x_2$ ranges. However, the statistical samples required for binning in all three variables would be far larger than are available at the LHC. To obtain measurements with small statistical errors we must integrate over two variables, namely $y$ and either $s$ or $\hat{m}_T$, and bin in the third variable. Fortunately, even though this involves convolution with the PDFs, many of the good qualities of the partonic relations discussed above survive to $d\sigma/d\hat{m}_T$ and $d\sigma/d\hat{s}$.

In our study of $pp \rightarrow V_1^0 V_2^0$ beyond LO in section 2.4, we will focus on $d\sigma/d\hat{m}_T$. However, our immediate goal in the remainder of section 2.3 is to explain heuristically how the ratios of eq. (2.2.2) behave, and to point out their most striking features. In this regard it is most useful to work with the variable $\hat{s} = m_{VV}^2$. The $\hat{m}_T$ and $y$ integrals split cleanly as separate functions of $\hat{s}$; see eq. (2.3.41) below. This feature makes formulas look simpler and permits simple heuristic arguments. Typically the features seen in $d\sigma/d\hat{m}_T$ are nearly the same as those seen in $d\sigma/dm_{VV}$, and moreover survive largely intact to NLO. We will see this for neutral diboson production later.

Of course the above-mentioned separation of $\hat{m}_T$ and $y$ integrals is only formal; it ceases to hold, even at LO, when realistic kinematic cuts are included. Such cuts are always necessary when photons are involved, since production rates diverge as
\( p_T^\gamma \to 0 \). Thus we must introduce a lower bound \((\tilde{m}_T)_{\text{min}}\) when integrating over \(\tilde{m}_T\) in eq. (2.3.37) to compute an observable rate. In section 2.4 below we bin with respect to \(\tilde{m}_T\), beginning at 200 GeV, so this requirement is automatically satisfied there. But in our heuristic LO discussion, where we bin with respect to \(m_{VV}\), we achieve this goal by imposing a cut on pseudorapidity

\[ |\eta(V)| < 1.5 \] (2.3.40)

for each final state boson \(V\); this cut renders the LO cross sections finite. This will not impact our heuristic reasoning but does play a role in the plots shown.

### 2.3.5 Ratio observables

We now discuss the ratio observables of eq. (2.2.2), already mentioned in section 2.2. We will present precise LO results in figures, and we will use schematic or approximate equations to understand the results. In this and following sections, all results are for a 13 TeV \(pp\) collider, and are obtained using MCFM 6.8 [102, 105]. The plots of our ratios are given for diboson cross sections without decays and do not include \(Z\) or \(W\) branching fractions to leptons.

For \(V_1^0 V_2^0 = \gamma \gamma, Z \gamma, ZZ\) we found that all the partonic cross sections are forward-backward symmetric and proportional to the kinematic function \(|a_1|^2\). For each of
<table>
<thead>
<tr>
<th>$V_1^0 V_2^0$</th>
<th>$C_{12}^u \cdot 10^5$</th>
<th>$C_{12}^d \cdot 10^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma\gamma$</td>
<td>1.2</td>
<td>0.7</td>
</tr>
<tr>
<td>$Z\gamma$</td>
<td>2.2</td>
<td>0.7</td>
</tr>
<tr>
<td>$ZZ$</td>
<td>1.6</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Table 2.1: The values of $C_{12}^q$ relevant for the $R_1$ ratios.

these processes, schematically,$^9$

$$
\frac{d\sigma_S}{d\hat{s}} (pp \to V_1^0 V_2^0) \sim \frac{\sum_q C_{12}^q \mathcal{L}^{S}_{qq}(\hat{s})}{\hat{s}^2} \int_{\sqrt{\hat{s}/2}}^{\sqrt{\hat{s}/2}} d\tilde{m}_T \left| \frac{d\tilde{t}}{d\tilde{m}_T} \right| |a_1|^2, \tag{2.341}
$$

where the $C_{12}^q$s were defined in eqs. (2.3.18)–(2.3.20). Note the numerator of the prefactor is a weighted parton luminosity, with the PDFs weighted by process-dependent couplings and charges. Our observable $R_{1a}$ then satisfies

$$
R_{1a}(\hat{s}) \equiv \left[ \frac{\sigma_S(pp \to Z\gamma)}{\sigma_S(pp \to \gamma\gamma)} \right] \sim \frac{\sum_q C_{12}^q \mathcal{L}^{S}_{qq}(\hat{s})}{\sum_q C_{12}^q \mathcal{L}^{S}_{qq}(\hat{s})}, \tag{2.342}
$$

with similar relations for $R_{1b} = \sigma_S(ZZ)/\sigma_S(\gamma\gamma)$ and $R_{1c} = \sigma_S(ZZ)/\sigma_S(Z\gamma)$.

One can then get a rough estimate for the $R_1$ ratios by using table 2.1 and applying the very crude relation $\mathcal{L}^{S}_{uu} \sim 2 \mathcal{L}^{S}_{dd}$. The small values of $C_{12}^u$ imply that $u\bar{u}$ initial states matter most for $R_{1a}$, and the parton luminosities largely cancel. We may therefore estimate $R_{1a} \sim C_{Z\gamma}^u/C_{\gamma\gamma}^u \sim 1.8$. Including $C_{12}^d$ and the crude relation among parton luminosities, the estimate increases to 2.1. This estimate is very good, as we can see by looking at the actual LO $R_{1a}$ ratio in figure 2.3. For ZZ, however, both $u\bar{u}$

---

$^9$The lower limit of integration over $\tilde{m}_T$ depends on the pseudorapidity cut imposed at $\eta_{cut} = 1.5$. In the $m_Z \to 0$ limit, $(\tilde{m}_T)_{\text{min}} = \sqrt{\hat{s}}/(2 \cosh \eta_{cut})$. The limits of integration over $y$ in $\mathcal{L}^{S}_{qq}$ also depend on $\tilde{m}_T$, a point we can ignore for the heuristic arguments presented here.
Figure 2.3: The $R_1$ ratios of $V_1^0V_2^0$ cross sections at LO, computed in MCFM at a $pp$ collider with $\sqrt{s} = 13$ TeV. A pseudorapidity cut of $|\eta(V)| < 1.5$ is imposed. These curves are determined almost entirely by ratios of parton luminosities, weighted by SM couplings.

and $d\bar{d}$ initial states are important. Although the similarly crude estimates $R_{1b} \sim 2.6$ and $R_{1c} \sim 1.3$ work quite well in the 1–2 TeV range, they are somewhat too small at low $\hat{s}$ because$^{10}$ $\mathcal{L}_{u\bar{u}}^S < 2\mathcal{L}_{d\bar{d}}^S$ for $\sqrt{\hat{s}} \ll 1$ TeV. We will see later that NLO QCD makes only minor corrections to these ratios, especially at high energy.

Next, we turn to the observables relating $W^+V^0$ and $W^-V^0$. We know from eq. (2.3.22) that the partonic cross sections $d\hat{\sigma}_S(W^+V^0)$ and $d\hat{\sigma}_S(W^-V^0)$ are identical. This leads to the following formula for the observable “charge asymmetry”,

$$C_{2a}(\hat{s}) \equiv \frac{\sigma_S(W^+\gamma)}{\sigma_S(W^-\gamma)} \sim \frac{\sum_{q_u,q_d} |V_{q_u,q_d}|^2 \mathcal{L}_{q_u,q_d}^S}{\sum_{q_u,q_d} |V_{q_u,q_d}|^2 \mathcal{L}_{q_u,q_d}^S},$$

written as a ratio of weighted parton luminosities, with $V_{ij}$ the CKM matrix. The

$^{10}$Effects from the $Z$ mass, neglected in these estimates, are indeed small, reaching only 3–6% for $\sqrt{\hat{s}} \sim 500$ GeV.
Figure 2.4: (Left) The $C_2$ charge ratios at LO, which go roughly like $f_u/f_d$ and are identical for $W\gamma$ and $WZ$. (Right) The $D_2$ variables, also identical for $W\gamma, WZ$. These forward-backward asymmetric charge ratios have a similar dependence on the PDFs, complicated by sign($y$) in the asymmetric parton luminosity which results in $|D_2| > C_2$.

The same result holds for $C_{2b} = \sigma_S(W^+Z)/\sigma_S(W^-Z)$. To derive an expectation for the magnitude and slope of these $C_2$ observables, we use the fact that $W^+V^0$ and $W^-V^0$ are produced predominantly at LO by $ud$ and $d\bar{u}$, respectively. Then we have roughly that $C_2 \sim \mathcal{L}_u^S/\mathcal{L}_{d\bar{u}}^S \sim f_u/f_d$, which has a magnitude of order 2, grows with energy, and is identical for $W\gamma$ and $WZ$ with negligible mass corrections. These expectations are confirmed in figure 2.4.

Similarly, because $d\tilde{\sigma}_A(W^+V^0)$ and $d\tilde{\sigma}_A(W^-V^0)$ are equal in magnitude and opposite in sign (see eq. (2.3.23)), we define

$$D_{2a}(s) \equiv \left[ \frac{\sigma_A(W^+\gamma)}{\sigma_A(W^-\gamma)} \right]_s \sim \frac{\sum_{q_u,q_d} |V_{q_u,q_d}|^2 \mathcal{L}_{d\bar{u}}^A}{-\sum_{q_u,q_d} |V_{q_u,q_d}|^2 \mathcal{L}_{d\bar{u}}^A} \sim -\frac{\mathcal{L}_{u}^A}{\mathcal{L}_{d\bar{u}}^A}. \quad (2.3.44)$$

An identical result, with negligible mass corrections, holds for the $WZ$ processes in $D_{2b}$. As we can see in figure 2.4, $D_2$ has a similar shape to $C_2$, but with opposite
sign and somewhat larger magnitude. This can be understood by recalling $\mathcal{L}^A_{q\bar{q}} = \int dy \, \text{sign}(y) \, 2 f_q(x_1) f_{\bar{q}}(x_2)$. If the $y < 0$ portion of the integral were zero, then we would have $|D_2| = C_2$. Instead, this portion is small, negative, and nearly identical for $\mathcal{L}^A_{ud}$ and $\mathcal{L}^A_{du}$. The fact that $|D_2|$ is fractionally larger than $C_2$ is merely a consequence of the inequality $(a - \epsilon)/(b - \epsilon) > a/b$ for $a > b > \epsilon > 0$.

Now we consider the observables that compare $W^{\pm \gamma}$ to $W^{\pm Z}$. Both $\sigma_A(W^{+\gamma})$ and $\sigma_A(W^{+Z})$ depend on the same weighted parton luminosity, which appears as the numerator of eq. (2.3.44). The antisymmetric partonic cross sections are equal in magnitude, opposite in sign, and proportional to $(a_1a_3)$. Everything thus cancels out of their ratio, leaving

$$A_2^+(\hat{s}) \equiv \left[ \frac{\sigma_A(W^{+\gamma})}{\sigma_A(W^{+Z})} \right]_{\hat{s}} \approx -1 . \quad (2.3.45)$$

As seen in figure 2.5, this ratio differs from $-1$ at low $\hat{s}$ due to few-percent $m_Z^2/\hat{s}$ effects. The same holds for the $W^{-V^0}$ processes in $A_2^-$. Since the PDFs are absent, these ratio observables can be computed with relatively low theoretical uncertainty. It is most unfortunate that these ratios have the largest statistical errors, as we will see in section 2.3.6.

As we discussed at the end of section 2.3.3.2, we naively expect

$$R_2^+(\hat{s}) \equiv \left[ \frac{\sigma_S(W^{+\gamma})}{\sigma_S(W^{+Z})} \right]_{\hat{s}} \sim \tan^2 \theta_W \approx 0.29 . \quad (2.3.46)$$

\footnote{In addition to the mass corrections to $(a_1a_3)$ given in Appendix A.1, the Jacobian $|dt/d\hat{m}_Z^2|$ and the limits of integration also have mass dependence that differs in numerator and denominator.}
Figure 2.5: (Left) The $R_2$ ratios, identical for $W^+, W^-$. These ratios are nearly $\tan^2(\theta_W)$, due to the coefficients of $|a_3|^2$ in the partonic rates. (Right) The $A_2$ ratios at LO, also identical for $W^+$ and $W^-$. These equal $-1$ because partonic forward-backward asymmetries are equal and opposite for $W\gamma$ and $WZ$, which depend on the same PDFs.

The one subtlety is the radiation zero in $a_3$ at $\theta = \pi/2$, which is potentially important because this is the region of phase space where $d\hat{\sigma}_S/d\hat{m}_T$ peaks (due to the Jacobian $|dt/d\hat{m}_T|$). However, as seen in figure 2.5, the above estimate is a good one. The reason is a combination of two pieces of good fortune. The first is that the ratio of the partonic amplitudes everywhere lies between 0.29 and 0.19. Since $|a_1|^2 = 2$ and $|a_\phi|^2 = 1/16$ at $\theta = \pi/2$, we see from eqs. (2.3.24) and (2.3.25) that

$$d\hat{\sigma}_S(u\bar{d} \to W^+\gamma) \propto s_W^2 Y_L^2, \quad d\hat{\sigma}_S(u\bar{d} \to W^+Z) \propto s_W^2 t_W^2 Y_L^2 + \frac{1}{32},$$

(2.3.47)

which means $d\hat{\sigma}_S(W^+\gamma)/d\hat{\sigma}_S(W^+Z) \to 0.19$ there. The second is that the coefficients of $|a_1|^2$ and $|a_\phi|^2$ are so small that $|a_3|^2$ is numerically very important despite its radiation zero.

This last statement is not true for $W^-W^+$; from eq. (2.3.28), the relative coefficient
of $|a_1|^2$ is 1/32, vs. $Y_L^2/8$ in $W\gamma$. Consequently $d\sigma_S(W^-W^+)$ is dominated by the
singlet term, making it roughly proportional to $d\sigma_S(V_1^0 V_2^0) \sim |a_1|^2$. This leads us to
consider ratios such as

$$R_{3a}(\hat{s}) \equiv \left[ \frac{\sigma_S(W^-W^+)}{\sigma_S(\gamma\gamma)} \right]_{\hat{s}},$$

(2.3.48)

and similarly $R_{3b} = \sigma_S(W^-W^+)/\sigma_S(Z\gamma)$ and $R_{3c} = \sigma_S(W^-W^+)/\sigma_S(ZZ)$. These
possibilities are displayed in figure 2.6. We can estimate their magnitudes just as we
did for the $R_1$ ratios above. Comparing the coefficients of $|a_1|^2$ in eqs. (2.3.17) and
(2.3.28), referring to table 2.1, and using the crude relation $\mathcal{L}_{\bar{u}u}^S \sim 2 \mathcal{L}_{dd}^S$, we get an
estimate

$$R_{3a} \sim \frac{1}{16} \left( \frac{\mathcal{L}_{\bar{u}u}^S + \mathcal{L}_{dd}^S}{s_W^4 Q_u^4 \mathcal{L}_{\bar{u}u}^S} \right) \sim 10.$$  

(2.3.49)

Similar estimates for $R_{3b}$ and $R_{3c}$ then follow from the $R_1$ ratios in figure 2.3.

Although these estimates are not wildly off, they do come up somewhat short, even
after allowing for $\mathcal{L}_{\bar{u}u}^S < 2 \mathcal{L}_{dd}^S$ at low $\hat{s}$. This is because we cannot actually ignore the
$|a_3|^2$ contribution to $\sigma(W^-W^+)$, which makes up about 20% of the total cross section.
Because of this, one may be led to include some admixture of $\sigma_S(W^\pm V^0)$ in the de-
nominators of the $R_3$ ratios. We leave it to further study to decide which admixture
would have the most desirable properties at NLO.

Finally, we turn to ratios involving $\sigma_A(W^-W^+)$. As we saw earlier, the leading or-
der partonic asymmetry in $W^-W^+$ is proportional to $(a_1a_3)$, as was the case for $W^\pm \gamma$
and $W^\pm Z$ (but see footnote 7). We therefore expect that a ratio of $d\sigma_A(W^-W^+)$ to


Figure 2.6: (Left) Possibly useful $R_3$ variables involving $W^{-}W^{+}$. Ratios are taken with $V_1^0V_2^0$ processes because $W^{-}W^{+}$ is dominantly produced as an $SU(2)$-singlet at LO. (Right) Possible $A_3$ variables involving forward-backward asymmetric $W^{-}W^{+}$ production. A property of the PDFs explains the flatness of the lower curve.

any linear combination of the $d\sigma_A(W^\pm V^0)$ is given by a ratio of parton luminosities weighted by SM coefficients. The asymmetries in $W^\pm Z$ suffer from low statistics, so we consider linear combinations of $d\sigma_A(W^+\gamma)$ and $d\sigma_A(W^-\gamma)$:

$$A_3(\hat{s}) \equiv \left[ \frac{\sigma_A(W^{-}W^{+})}{a \sigma_A(W^+\gamma) + b \sigma_A(W^-\gamma)} \right] \sim \left( a \mathcal{L}_{u \bar{u}}^A - b \mathcal{L}_{d \bar{d}}^A \right) \frac{\mathcal{L}_{ud}^A}{4 |V_{ud}|^2 s_W Y_L} \frac{\mathcal{L}_{ud}^A}{a \mathcal{L}_{u \bar{u}}^A - b \mathcal{L}_{d \bar{d}}^A}. \quad (2.3.50)$$

It is an interesting non-obvious feature of the PDFs that, as functions of $\hat{s}$ in the kinematic region of interest,

$$\mathcal{L}_{u \bar{u}}^A \propto \mathcal{L}_{d \bar{d}}^A \propto \mathcal{L}_{ud}^A + \mathcal{L}_{d \bar{d}}^A; \quad (2.3.51)$$

the first (second) relation holds at the 2% (15%) level. This suggests the use of $a = -b = 1$, which has the further advantage of minimizing the relative statistical uncertainty in the denominator of eq. (2.3.50). Whether this is the ideal choice after NLO
corrections are included remains to be seen. We can see in figure 2.6 that, at LO, this choice leads to a much flatter and smaller ratio than the choice \( a = b = 1 \).

### 2.3.6 Limitations of finite statistics

Attractive as these ratios are, the reality of low cross sections means that many of these observables are not useful in the near term. In table 2.2 we show a rough estimate of the number of high-energy events \( (m_{VV} > 400 \text{ GeV}) \) expected for each process. We assume 300 fb\(^{-1} \) at \( \sqrt{s} = 13 \text{ TeV} \) and account for leptonic branching fractions of the \( Z \) and \( W \). We computed these numbers imposing a pseudorapidity cut \( |\eta(V)| < 1.5 \) on the bosons (as in table 2.3 below), and have separated events into “Forward” and “Backward” by the sign of \( y(\eta_1 - \eta_2) \) as described in section 2.3.4.

Any one of our ratios becomes interesting as a precision observable once its statistical uncertainty becomes of order 5–10\%, so that its exceptionally low theoretical errors become experimentally relevant. If such small uncertainties are possible for a particular ratio only by combining all events together into a single bin, e.g. using ratios of total cross sections with \( \tilde{m}_T > 200 \text{ GeV} \), then this measurement is likely to be useful only for testing methods for SM predictions, since it will be sensitive mainly to physics only up to the \( \sqrt{s} \sim 400 \text{ GeV} \) range. However, more can be done once the events can be divided into multiple bins of varying width, each with statistical uncertainty of order 5–10\%, as in figure 2.1. In this case the lower bins serve as a test of
<table>
<thead>
<tr>
<th>$V_1V_2$</th>
<th>$N_f + N_b$</th>
<th>$N_f - N_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma\gamma$</td>
<td>12000</td>
<td>0</td>
</tr>
<tr>
<td>$Z\gamma$</td>
<td>2000</td>
<td>0</td>
</tr>
<tr>
<td>$ZZ$</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td>$W^+\gamma$</td>
<td>3300</td>
<td>$-500$</td>
</tr>
<tr>
<td>$W^-\gamma$</td>
<td>2100</td>
<td>220</td>
</tr>
<tr>
<td>$W^+Z$</td>
<td>790</td>
<td>33</td>
</tr>
<tr>
<td>$W^-Z$</td>
<td>520</td>
<td>$-16$</td>
</tr>
<tr>
<td>$W^-W^+$</td>
<td>9500</td>
<td>$-430$</td>
</tr>
</tbody>
</table>

Table 2.2: At LO, the number of events with $\sqrt{s} = m_{VV} > (400 \text{ GeV})^2$ assuming 300 fb$^{-1}$ and leptonic decays of $W$ and $Z$. $N_f$ and $N_b$ indicate forward- and backward events. These numbers increase by a factor of order 1.5–2 at NLO, but are reduced by a comparable amount when using the variable $\bar{m}_T$ instead of $m_{VV}$.

The predictive techniques, while the higher ones are useful for other purposes, including searches for BSM phenomena and tests of important EW corrections that grow with energy and do not entirely cancel in these ratios.

As we saw in figure 2.1 of section 2.2, the ratio $R_{1a}$ permits 6 bins at 300 fb$^{-1}$ with 6% statistical uncertainties. At this integrated luminosity, the other variables that allow multiple bins with $\sim$5% uncertainties are $C_{2a}$ and $R_3$, as one can see using table 2.2. Meanwhile $R_{12}^\pm$, $C_{2b}$, and $D_{2a}$ allow for a single bin.

The situation will improve at 3000 fb$^{-1}$, though the high pileup environment may lead to some loss of statistics. If we simply assume the total rate increases by a factor of 10 without significant losses, we find that in addition to the above six variables, the variables $R_{3b}$, $R_{1c}$ and $A_3$ also permit multiple bins. The $A_2^+$ ratio can be used
in a single bin. The two variables $A_2^-$ and $D_{2b}$ involving $\sigma_A(W^- Z)$ are too small to measure.

It may prove useful to improve statistics slightly by combining observables predicted to be equal within the SM. For instance, one could replace $R_2^+$ and $R_2$ with

$$R_2^0 = \frac{\sigma_S(W^+ Z) + \sigma_S(W^- Z)}{\sigma_S(W^+\gamma) + \sigma_S(W^-\gamma)}. \quad (2.3.52)$$

Similar combinations would assist with $A_2^+$ and $A_2^-$ (see figure 2.5), $C_{2a}$ and $C_{2b}$, and $D_{2a}$ and $D_{2b}$ (see figure 2.4).

### 2.4 Beyond leading order for $\gamma\gamma$, $Z\gamma$, $ZZ$

In section 2.3.3 we saw that the differential LO partonic cross sections for $V_1^0 V_2^0 = \gamma\gamma$, $Z\gamma$, $ZZ$ are all proportional to the same function $|a_1|^2$, up to $m_Z^2/p_T^2$ effects (provided in Appendix A.1). Consequently, at high energy, the ratios of these partonic cross sections are given by constants of the SM. Since the up quark PDF dominates $\gamma\gamma$ and largely dominates $Z\gamma$, the hadronic ratio $R_{1a}$ is approximately constant and equal to a simple partonic ratio. Although the PDFs have a greater effect on the hadronic cross sections for $R_{1b}$ and $R_{1c}$, these two observables still vary rather slowly with $\sqrt{s} = m_{VV}$, with easily understandable values, as we saw in figure 2.3.

Beyond LO, we will study the $R_1$ ratios differentially with respect to $\bar{m}_T$, the aver-
Figure 2.7: The $R_1$ ratios at LO binned in $\tilde{m}_T$.

age $m_T$ of the two vector bosons, eq. (2.3.34). The LO ratios in this variable are given in figure 2.7. Comparing with figure 2.3, one can see that the LO ratios as functions of $\tilde{m}_T$ and as functions of $m_{VV}/2$ are quite similar. This is because the hadronic cross section for a given $\tilde{m}_T$ is dominated by the region with $\sqrt{s} \sim 2\tilde{m}_T$.

The fully-differential cross-sections for the diboson processes have been known for quite some time [42, 183, 185, 267, 280–283, 297]. In this and following sections, all calculations are carried out using MCFM 6.8 [102, 105, 159], except for an NNLO real emission study which used MadGraph 2.3.0 [34]. Renormalization and factorization scales $\mu_R, \mu_F$ are chosen at $m_{VV}$ except when otherwise specified. We use MSTW 2008 NLO [LO] PDFs [265] for all NLO [LO] calculations and for $O(\alpha_S^3) [O(\alpha_S^2)]$ $gg \to V_1^0V_2^0$ calculations. Our cuts on the bosons are presented in table 2.3. See section 2.5 for cuts on their decay products.
2.4.1 Choices of observable and of cuts

We begin with a discussion of our cuts and our observable. It is important to choose these carefully in order to avoid large NLO and NNLO corrections to our ratios, and associated large uncertainties.

We will discuss certain experimental realities in section 2.5, but for now we neglect $Z$ decay and impose cuts on the vector bosons and on any jets,$^{12}$ as in table 2.3. (Our cuts on leptons in $Z \rightarrow \ell^+\ell^-$ are given in table 2.5 of section 2.5.2.) In our discussion we will have at most one jet and so for us $H_T$ is simply the $p_T$ of that jet, but it is important that $H_T$ be the variable used at higher jet multiplicity, not maximum jet $p_T$. This cut ensures that multiple jets with $p_T$ just below our cuts cannot combine together on one side of the event and force the two bosons to be close in angle, or allow one boson to be soft relative to the QCD activity. Either of these effects would allow events that are far in phase space from the LO kinematics to enter the measurement, and potentially cause large corrections and failures of cancellations in our ratios. Note also that we choose identical kinematic cuts for $Z$ and $\gamma$, which we supplement in section 2.5 when being more experimentally realistic. We discuss angular isolation of the bosons in section 2.4.2 and section 2.5.1.

A variety of problems can arise that can invalidate or destabilize fixed-order calcu-

---

$^{12}$We will refer to all final-state colored partons, for brevity only, as “jets”. We do not include showering and hadronization in our study, but we expect these to have small effects, since we impose cuts on our observables to avoid regions where resummation plays an important role.
<table>
<thead>
<tr>
<th>Kinematic Cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
</tr>
<tr>
<td>$p_T(V_2) &gt; \frac{1}{2} p_T(V_1)$</td>
</tr>
<tr>
<td>$H_T = \sum_{jets}</td>
</tr>
</tbody>
</table>

Table 2.3: Kinematic cuts imposed on vector bosons $V_i$ and on the jets $j$ from real emission at NLO. In our calculations we work only to single real emission so $H_T$ is simply $p_T^j$, but the use of an $H_T$ cut is important at higher orders. We define $V_1, V_2$ by $p_T(V_1) > p_T(V_2)$. Isolation requirements and cuts on decay products are described in section 2.5.

Our cuts, which allow the vector bosons to have unequal $p_T$, but require both bosons have substantially higher $p_T$ than any jet from real emission, are chosen to avoid them. Note also that our cuts generally scale with the overall average $p_T$, and roughly with our observable $\bar{m}_T$.

One issue we must avoid is large logarithms. The fairly loose cut on additional hadronic activity, $H_T < \frac{1}{2} p_T(V_2)$, means that logarithms of $p_T(V)/H_{T,\text{min}}$ never become so large as to require jet veto resummation [46, 49, 303]. But because our cuts scale with the average $p_T$, we also avoid large logarithms of $p_T(V_1)/p_T(V_2)$, which (in combination with a large $gg$ parton luminosity) could have led to very large corrections [291]. Simultaneously,\textsuperscript{13} asymmetric cuts on the bosons avoid logarithms of $p_T(V)/\Delta$, where $\Delta = p_T^{\text{cut}}(V_1) - p_T^{\text{cut}}(V_2)$; these logarithms, which arise from soft gluon

\textsuperscript{13}We thank Z. Bern for alerting us to possible subtleties with these cuts and specifically to ref. [186].
emission, were first identified in ref. [186] and resummed in ref. [45]. Meanwhile our observable itself, $\hat{m}_T$, does not appear in large logarithms and requires no resummation.

Other effects can enhance the size of fixed-order terms relative to naive expectations. For instance, if radiative corrections are allowed to populate phase space at lower $\sqrt{s}$ than is accessible at tree level, the formally NLO calculation carries de facto LO scale uncertainties. This does not happen with our cuts and observable; all bins in $\hat{m}_T$ are dominated by the LO contribution.

Another common issue with $q\bar{q}$ processes is the opening of new channels with large parton luminosities at higher orders. At NLO, we have the new channel $gg \rightarrow qV_1^0V_2^0$, but our cuts mitigate the $K$ factors, making them of order 1.5. Moreover, these $K$ factors are nearly process independent and largely cancel in our ratios. At NNLO, we have the new channel $gg \rightarrow V_1^0V_2^0$, which is substantial and process dependent; we include it in our calculation. Also at NNLO is the new channel $qq \rightarrow qqV_1^0V_2^0$, which is process dependent and potentially large for valence quarks. We estimate that with our cuts, (which avoid any large logarithmic enhancements,) this process is subleading; we do not evaluate it but include it in our uncertainty estimates.

We must also avoid situations where higher-order matrix elements (at a particular jet multiplicity) are enhanced relative to LO matrix elements dressed with soft and collinear factors (at the same multiplicity). One way this can happen is if an additional jet emission can make a threshold or resonance accessible that was inaccessible
at lower order. This can occur in QCD corrections to the $Z\gamma$ process, via radiative $Z$
decays $Z \to \ell^+\ell^-\gamma$. Simply because we take $\bar{m}_T > 200$ GeV, this is irrelevant at LO,
and our $H_T$ cut assures this does not arise at any order in $\alpha_S$.

A further potential problem can appear if a radiative emission can significantly de-
crease an internal propagator’s virtuality compared to the analogous propagator in
the LO process, thus enhancing the amplitude. (Strictly speaking, this way of stating
things is not gauge invariant, but the enhancement itself clearly is.) With our cuts
and observable, this too does not occur.

\subsection{NLO QCD corrections}

For our observables and with our choice of cuts, virtual and real QCD corrections to
$d\hat{\sigma}(q\bar{q} \to V_1^0 V_2^0)$ are largely proportional to the LO values. Consequently the $R_1$ ratios
receive only small NLO QCD corrections in most regions of phase space. The excep-
tion is in the region where a final-state quark is nearly collinear with a vector boson;
this region is enhanced for photons by large logs from collinear emission, whereas for
$Z$s the logarithmic enhancement is cut off by $m_Z$. More specifically, for $Z$ emission
the quark propagator in figure 2.8 is bounded from above by $1/m_Z^2$, while the pho-
ton’s collinear singularity at low $m_{q\gamma}$ must be absorbed into a non-perturbative frag-
mentation function, or evaded through an angle-dependent energy isolation cut that
avoids generating soft divergences at higher order. This fundamental difference be-
Figure 2.8: The regime in which $V$ and $q$ are nearly collinear in the final state, the source of a significant difference between photon- and $Z$-rates.

tween $Z$ and $\gamma$ cannot be removed experimentally, and gives a significant NLO shift to the $R_1$ ratios at low $\bar{m}_T$.

The collinear-$\gamma q$ singularity can be dealt with using the smooth-cone isolation method of Frixione [184]. (While theoretically elegant, this method is not practical; we will employ a more experimentally realistic version of Frixione isolation, and discuss the uncertainties inherent in its use, in section 2.5.1.) In this method, one chooses two parameters $\delta, \epsilon$ and requires that in any cone of radius $R < \delta$, the hadronic activity is bounded by a function that goes smoothly to zero as $R \to 0$; in particular\(^{14}\)

$$
\sum_{h \in R} p_T^h < p_T(V) \quad \mathcal{I}(R; \epsilon, \delta) \quad \text{for all } R < \delta ,
$$

(2.4.1)

$$
\mathcal{I}(R; \epsilon, \delta) = \epsilon \left( \frac{1 - \cos R}{1 - \cos \delta} \right).
$$

(2.4.2)

Here the sum is over all hadrons $h$ within a cone of radius $R$ around the boson.

That the $R_1$ ratios remain unchanged outside the collinear regime may be seen by applying the Frixione method with extreme parameters $\left( \delta, \epsilon \right) = (1.2, 0.2)$. This choice

\(^{14}\)Frixione included a third parameter $n$ as an exponent on the trigonometric function here; we have chosen $n = 1$.  

54
Figure 2.9: The $V_1^0 V_2^0$ cross section at NLO, shown relative to the LO rates. At left, the collinear region was removed by a very strict smooth-cone isolation cut $(\delta, \epsilon) = (1.2, 0.2)$ applied to both $\gamma$s and $Z$s. All 3 processes receive identical NLO corrections, thus leaving the ratios invariant. At right, with a reasonable isolation cut $(\delta, \epsilon) = (0.4, 0.5)$ the NLO corrections differ significantly among the processes at low energies.

largely removes the collinear region. Here (but see below) we apply isolation both to photons and $Z$s, to maintain as much congruence as possible. At left in figure 2.9, we see that the $K$ factors are then almost identical for the three $V_1^0 V_2^0$ processes, and so the $R_1$ ratios at NLO are the same as at LO.

However, as seen at right in the same figure, when the collinear region is restored by using more reasonable smooth-cone parameters $(\delta, \epsilon) = (0.4, 0.5)$, there is a significant splitting in the $K$ factors at low $m_T$, where the $Z$ mass is particularly relevant, and thus a shift in the $R_1$ ratios away from their LO values. Note that the splitting of $\gamma\gamma$ from $ZZ$ is roughly double that of $\gamma\gamma$ from $Z\gamma$, so the effect of the collinear regime is largest on $R_{1b}$.

In all results beyond this point we use $(\delta, \epsilon) = (0.4, 0.5)$, with appropriate practical
modifications discussed in section 2.5.1. For this choice, and for the range of \( \bar{m}_T \) that is relevant for the LHC, we find it unnecessary to impose isolation on \( Zs \), for the following reasons. At low \( \bar{m}_T \) the Frixione cut removes a region where the amplitude for \( Z \) emission is not enhanced. Meanwhile at larger \( \bar{m}_T \) the falling \( qg \) parton luminosity makes the collinear region less important even for photons, an effect seen at right in figure 2.9, and also tends to favor the region of low \( p_T^q/p_T^Z \), which is not removed by the Frixione cut. Altogether this reduces the impact of Frixione isolation on \( Zs \) to the percent level, relative to the total differential cross section. Therefore, in what follows below and in our final results, we impose isolation only on photons, not on \( Zs \), and believe it is safe for the LHC experiments to do the same without negatively impacting the ratios. At a higher-energy collider this would need to be revisited.

With these Frixione parameters, our lowest \( \bar{m}_T \) bin sees a downward shift of \( R_{1a} = \sigma(Z\gamma)/\sigma(\gamma\gamma) \) by 15%, of \( R_{1b} = \sigma(ZZ)/\sigma(\gamma\gamma) \) by 25%, and of \( R_{1c} = \sigma(ZZ)/\sigma(Z\gamma) \) by 12% relative to the LO values. In higher bins, the effect of the collinear regime is muted as the \( qg \) parton luminosity falls and the difference between photon and \( Z \) amplitudes decreases.

It is instructive to understand why the NLO corrections to the \( R_1 \) ratios are so small outside of the collinear region. The point is that most logarithmically-enhanced corrections are themselves proportional to the LO process, for reasons that even extend to many regions of phase space that are not log-enhanced. For instance, in the NLO process \( q\bar{q} \rightarrow V_1^0V_2^0g \), our cuts are inclusive in the initial state radiation (ISR)
region of phase space, where the final-state gluon is collinear with the initial partons. Consequently a fixed-order calculation is a reliable guide, and the NLO diagrams that appear are the same for all three processes. Thus no large process-dependent corrections arise, and the $R_1$ ratios are hardly affected. Meanwhile emissions of hard gluons are suppressed by our jet cuts.

Similarly, for the ISR region of $qg \rightarrow V_1^0V_2^0q$, the ratios are little changed, for two reasons. First, the partonic cross section near this singular region displays a factorization into the tree-level cross section and a universal factor that is absorbed into the definition of the PDFs. Second, the replacement of an anti-quark PDF with a gluon PDF has a small impact, because $\bar{u}$ and $\bar{d}$ PDFs are similar. We may see this heuristically by writing $f_\bar{q} = \frac{1}{2}(f_{\bar{u}} + f_{\bar{d}})$ and $\bar{\delta} = \frac{1}{2}(f_{\bar{u}} - f_{\bar{d}})$, and noting the $qg$ integrand is roughly proportional to

$$[f_u(x_1)d\hat{\sigma}_{uu}^{LO} + f_d(x_1)d\hat{\sigma}_{dd}^{LO}] [f_g(x_2/z)P_{q\leftarrow g}(z)]$$

(2.4.3)

while the tree-level process has integrand

$$[f_u(x_1)d\hat{\sigma}_{uu}^{LO} + f_d(x_1)d\hat{\sigma}_{dd}^{LO}] f_\bar{q}(x_2) + O(\bar{\delta}) .$$

(2.4.4)

Here $P_{q\leftarrow g}$ is the gluon-to-quark splitting function, and we have ignored small contributions from subdominant initial states. Since $\bar{\delta} \ll f_\bar{q}$ in the relevant $x$ range, these integrands are proportional, so no large correction to the LO ratios is expected from the ISR region.


2.4.3 NNLO QCD corrections

Although NNLO calculations of diboson processes have been carried out for all processes except $WZ$ [108, 109, 194, 207–209], most of these are not yet accessible in public code. This limits our ability to refine our NLO results or to estimate the theoretical uncertainties from which they suffer. In this context, we take the following approach. On the one hand, we study in detail the largest known NNLO correction to our ratios, namely $gg \to V_1^{0}V_2^{0}$, which is large enough that it must be included, but fortunately is available publicly. On the other hand, we search for additional NNLO corrections that should affect our ratios, and make rough estimates of their size to see if they are important; if so we include them as a theoretical uncertainty.

We saw in figure 2.9 and eqs. (2.4.3)–(2.4.4) that many NLO corrections are common to all three $V_1^{0}V_2^{0}$ processes and cancel in the $R_1$ ratios. Similar logic would suggest that many NNLO QCD corrections are also common to the three processes and that, away from the collinear-$qV$ regions, new real contributions like $q\bar{q} \to V_1^{0}V_2^{0}gg$, or $gg \to V_1^{0}V_2^{0}gg$ are likely to cancel. But by looking carefully at the physical origin of various effects, we can also see where such cancellations will fail.

Before we do so, let us forestall an obvious question. Below, we will assume that many NNLO corrections cancel in ratios, and that the largest one that does not cancel comes from the $gg \to V_1^{0}V_2^{0}$ loop graph (as suggested in ref. [80]), which we will include explicitly below. One might question this assumption based on the existing
NNLO and near-NNLO literature, which suggests potentially large $K_{\text{NNLO/NLO}}$ factors (1.3 – 1.6), substantial process-dependence in these $K$ factors, and effects that can be much larger than the $gg \rightarrow V_1^0 V_2^0$ loop graph. How, then, can we possibly claim that NNLO corrections to our ratios could be brought under control, and further assume that even higher-order effects can be ignored?

Here one needs to look carefully at the details, which we do in Appendix A.3. The large $K_{\text{NNLO/NLO}}$ arise only in situations where the cuts on the bosons and jets are very different from our own, causing even the $K_{\text{NLO/LO}}$ factor to be much larger than the $\sim 1.5$ that we found above in figure 2.9. The process-dependent differences among the $K_{\text{NNLO/NLO}}$ factors also appear much smaller when one restricts to kinematic regions and observables similar to the ones we are considering. In those regions there is no clear indication that the $gg$ loop is not the main process-dependent effect. Thus there is no clear evidence against our assumptions, and even some mild (though hardly decisive) evidence in their favour. Let us note again that our choice of observable and of cuts appears to be crucial in this regard; many other observables and cuts would have larger NNLO corrections in ratios.

With that issue set aside, we now consider obvious sources of NNLO corrections that will not cancel in our ratios. Since the dominant NLO correction to the ratios, shown in the right-hand plot of figure 2.9, was from the collinear-$qV$ region, corrections to that region of phase space will not cancel. NNLO real and virtual corrections to this single-collinear effect will impact the ratios. However, we expect these to give
an order $\alpha_s$ adjustment to the splitting shown in figure 2.9, which puts them below 2%.

Another important contribution could come from the double-collinear region in $q\bar{q}, gg \rightarrow V_1^0 V_2^0 q\bar{q}$. This too is very small, despite the large NLO single-collinear correction. To see this, note the following. The reason that $qg \rightarrow V_1^0 V_2^0 q$ is so important is that $\mathcal{L}_{qg} \gg \mathcal{L}_{q\bar{q}}$, partially canceling the extra $\alpha_s$ at NLO. There is no corresponding enhancement for two independent collinear emissions. The double-collinear region at the next order should be thought of predominantly as $q\bar{q} \rightarrow q\bar{q}, gg \rightarrow q\bar{q}$, with double emission $q \rightarrow qV_1$ and $q \rightarrow \bar{q}V_2$. (Our cuts remove the region where both $V_1$ and $V_2$ radiate off a single quark.) For $q\bar{q} \rightarrow q\bar{q}$ the parton luminosity is the same as that arising at LO, so the $q\bar{q}$-initiated process is indeed suppressed by $O(\alpha_s^2) \sim 1\%$ compared to LO. Meanwhile, $\mathcal{L}_{gg}$ is comparable to or smaller than $\mathcal{L}_{qg}$ at the relevant energies; and furthermore $gg \rightarrow q\bar{q}$, which lacks a $t$-channel gluon, has a smaller partonic cross section than $qq \rightarrow q\bar{q}$ and $q\bar{q} \rightarrow q\bar{q}$. Altogether it appears the double-collinear regime shifts the ratios at the percent level or below.

A qualitatively new source of non-canceling corrections is from the opening of a new channel at NNLO, namely the (dominantly valence-quark) process $qq \rightarrow qqV_1^0 V_2^0$. When each of the two fermion lines emits one vector boson, the resulting contribution is generally no longer proportional to the LO $qq \rightarrow V_1^0 V_2^0$ process. Still, we estimate that the $qq$-initiated processes at NNLO correct the ratios by just a few percent. Our argument proceeds as follows. The process $qq \rightarrow qqV_1^0 V_2^0$ has a collinear
divergence near the beampipe and can only be defined by requiring both jets to have
$p_T$ greater than some minimum $p_T^{i,\text{min}}$. However, the divergence is proportional to
the LO $q\bar{q} \to V_1^0V_2^0$ process, and largely cancels in the ratios. Calculating the effect
on the ratios for different values of $p_T^{i,\text{min}}$ between 5 and 30 GeV, and extrapolating
$p_T^{i,\text{min}} \to 0$ by fitting to a falling exponential, we find shifts for $R_{1a}$ ($R_{1b}$) [$R_{1c}$] of 3%
(3.5%) [2.5%] or less. Consequently, although our estimates are crude and this source
of NNLO corrections may well be one of the largest on the $R_1$ ratios, it does not seem
to present issues that exceed our fiducial benchmark of 5–6% theoretical uncertainties.

Finally, the largest known NNLO correction to the $R_1$ ratios is from $gg \to V_1^0V_2^0$.
Fortunately, much is already known about this correction, which is separately gauge-
invariant and finite. It has been known for some time [144, 202] and can consistently
be combined with the NLO calculation on its own. As it gives the largest source of
NNLO corrections in most regions of phase space and has a different dependence on
EW quantum numbers than does the tree-level process, it has an important effect on
our ratio observables.

Because $u$- and $d$-type quarks contribute coherently in the loop, the formulas for
$gg \to V_1^0V_2^0$ are not proportional to the tree-level $q\bar{q} \to V_1^0V_2^0$ formulas. In fact
$gg \to w^3x$ is zero by $SU(2)$ conservation, and so $gg \to Z\gamma$ is relatively small com-
pared to $gg \to ZZ, \gamma\gamma$. In figure 2.10 the $gg$ contributions to the cross sections are
shown relative to the corresponding NLO differential cross sections; they represent a
13% (5%) [20%] correction for $\gamma\gamma$ ($Z\gamma$) [$ZZ$] at low $m_T$, though less at higher energies
Figure 2.10: (Left) Contribution from $gg \rightarrow V_1^0 V_2^0$ to the $V_1^0 V_2^0$ cross sections, expressed relative to the corresponding NLO cross section. We used $gg \rightarrow \gamma \gamma$ at $O(\alpha_s^3)$ to estimate $gg \rightarrow Z\gamma, ZZ$ at this order. (Right) The $R_1$ ratios, including the NLO and $gg \rightarrow V_1^0 V_2^0$ contributions.

where the gluon PDFs are smaller.

Partial cancellations still take place in our ratios. The observable $R_{1a}$ is shifted downward by as much as 7% from its NLO value at the lowest values of $\bar{m}_T$ we consider; however, this $gg$-shift is reduced at higher $\bar{m}_T$, quickly becoming of order 3%. Meanwhile $R_{1b}$ ($R_{1c}$) shifts up 7% (14%) at low $\bar{m}_T$; this $gg$-shift remains at the 6% (9%) level for moderate $\bar{m}_T$ before shrinking more rapidly to 3% (3%) at high $\bar{m}_T$.

Figure 2.10 displays the $R_1$ ratios including the $gg \rightarrow V_1^0 V_2^0$ channel along with the NLO contributions. This plot should be compared with figure 2.7, which shows the LO ratios. Notice that $R_{1a}$ is accidentally flatter than at LO, as a result of the above-mentioned corrections.

This plot of course depends on a choice of renormalization and factorization scales $\mu_R$ and $\mu_F$ used for the $gg \rightarrow V_1^0 V_2^0$ computation. For $gg \rightarrow \gamma \gamma$ the scale dependence
can be reduced because the dominant\(^{15}\) part of the \(O(\alpha_S^3)\) correction is known \([80]\).

For \(gg \to Z\gamma, ZZ\), we can use the fact that at NLO all three processes have a nearly universal \(\mu_R, \mu_F\) dependence for \(s \gg m_Z^2\). This is because (i) the three processes have the same \(\alpha_S\)-dependence and involve the same PDFs, (ii) the SM is anomaly free and so no new non-universal diagrams appear at \(O(\alpha_S^3)\), and (iii) the contribution of longitudinal Zs to \(gg \to ZZ\) is rather small \([202]\), of order 10–15%. Thus for reasonable values of \(\mu_R\) and \(\mu_F\),

\[
K_{gg} \equiv \frac{d\sigma_{(3)}(gg \to \gamma\gamma)}{d\sigma_{(2)}(gg \to \gamma\gamma)} \approx \frac{d\sigma_{(3)}(gg \to Z\gamma)}{d\sigma_{(2)}(gg \to Z\gamma)} \approx \frac{d\sigma_{(3)}(gg \to ZZ)}{d\sigma_{(2)}(gg \to ZZ)},
\]

where \(d\sigma_{(n)}\) marks the cross section calculated at order \(\alpha_S^n\). We can then use MCFM to compute the known \(O(\alpha_S^2)\) and \(O(\alpha_S^3)\) cross sections for \(gg \to \gamma\gamma\), thereby determining the \(O(\alpha_S^3)\) cross sections for the other processes to a fairly good approximation. For our central values we choose scales \(\mu_R = \mu_F = m_{\gamma\gamma}\) everywhere in eq. (2.4.5).\(^{16}\)

We show the values of \(K_{gg}\) in left panel of figure 2.11. Since the values of \(K_{gg}\) are large, one might wonder whether, as in \(gg \to h\), the \(O(\alpha_S^4)\) correction to \(gg \to V_1^0V_2^0\) could itself be quite large. However, unlike \(gg \to h\), where the NLO prediction ex-

\(^{15}\)In Appendix A.2 we argue that the terms neglected in ref. \([80]\) are indeed subleading. For \(gg \to ZZ\) a similar calculation appeared very recently \([106]\), as this paper was nearing completion.

\(^{16}\)We have observed, by direct comparison across our \(\bar{m}_T\) range, that the procedure just outlined is essentially identical to calculating the \(O(\alpha_S^3)\) cross sections for the three processes with scales \(\mu_R \sim 0.34 m_{VV}\) and \(\mu_F \sim 0.20 m_{VV}\). The fact that these are reasonable scales serves as a sanity check of our method.
Figure 2.11: (Left) The size of $O(\alpha_S^3)$ corrections $K_{gg}$ to $gg \to \gamma\gamma$ as a function of $m_T$, with scales set to $\mu_R = \mu_F = m_{\gamma\gamma}$ in numerator and denominator. This function allows us to estimate the $O(\alpha_S^3)$ cross section for $gg \to Z\gamma$ and $ZZ$. (Right) As a function of scale $\mu_R = \mu_F = \mu$, the $gg \to \gamma\gamma$ rate in the kinematic region $m_T > 200$ GeV, shown at $O(\alpha_S^2)$ and (with the partial calculation implemented in MCFM) at $O(\alpha_S^3)$. The cross sections are normalized with respect to $\sigma_0 \equiv \sigma(3)(gg \to \gamma\gamma, \mu = m_{VV})$.

ceeds the LO substantially at all $\mu$, the situation is milder here. As can be seen in the right panel of figure 2.11, which shows $gg \to \gamma\gamma$ at $O(\alpha_S^2)$ and $O(\alpha_S^3)$ with a variety of scale choices, the higher-order prediction turns over at small $\mu$, and above the turnover varies only slowly. We therefore expect $O(\alpha_s) \sim 10 - 20\%$ uncertainties on $gg \to \gamma\gamma$, and $\sim 1 - 2\%$ uncertainties on the $R_1$ variables, from the unknown $O(\alpha_S^4)$ terms. We will estimate uncertainties from this source in section 2.4.4 and find them consistent with this expectation.
Figure 2.12: The relative PDF uncertainty bands for the individual $V_1^0V_2^0$ cross sections (left) and the $R_1$ ratios (right). PDF variations of $gg \rightarrow V_1^0V_2^0$ are included. See text for more details.

2.4.4 Partial cancellation of PDF and scale uncertainties

Now we turn to standard sources for potential theoretical uncertainties: the PDFs and the choices of renormalization and factorization scales in QCD corrections. These show significant cancellations and become subleading compared to other uncertainties that we have already discussed.

The PDF uncertainties for the individual channels, and their reduced values for the ratios, are shown in figure 2.12. For the $R_{1\alpha}$ ratio the uncertainties are of order 1% and can be essentially ignored; as we saw in section 2.3.5, the parton luminosity $\mathcal{L}^S_{ui\bar{u}}$ dominates both numerator and denominator, so that PDF variations nearly cancel. For the others, the uncertainties are still significantly reduced, rising only to about 2% even up to $m_T \sim 1$ TeV.

These uncertainties were determined using MCFM 6.8. The $pp \rightarrow V_1^0V_2^0$ cross
sections are evaluated for the central ($S_0$) and all 20 pairs of error sets ($S_i^{\pm}$) of the MSTW 2008 PDF set [265]. With the cross sections $d\sigma(S)$, we use the prescription of ref. [265] to determine the PDF uncertainties on individual channels. The upper edge of the uncertainty band is calculated with

$$\Delta_+(d\sigma) = \sqrt{\sum_i \left( \max \left[ 0, d\sigma(S_i^+) - d\sigma(S_0), d\sigma(S_i^-) - d\sigma(S_0) \right] \right)^2},$$

(2.4.6)

while the lower edge is the same with “max” replaced with “min”.$^{17}$ Because the error sets of MSTW 2008 are eigenvectors of the covariance matrix, the PDF uncertainties for the ratios can then be obtained in a similar fashion.$^{18}$

All this is straightforward except for one subtlety. Since we do not have access to the $O(\alpha_s^3)$ calculation for $gg \rightarrow Z\gamma$ and $gg \rightarrow ZZ$, we obtain them by rearranging eq. (2.4.5) as

$$d\sigma_{(3)}(gg \rightarrow Z\gamma, \text{pdf}_1) \approx d\sigma_{(3)}(gg \rightarrow \gamma\gamma, \text{pdf}_1) \frac{d\sigma_{(2)}(gg \rightarrow Z\gamma, \text{pdf}_2)}{d\sigma_{(2)}(gg \rightarrow \gamma\gamma, \text{pdf}_2)},$$

(2.4.7)

where $d\sigma(\ldots; \text{pdf}_i)$ is the cross section evaluated for PDF set $S_i$. A similar expression holds for $gg \rightarrow ZZ$. Inaccuracies in this procedure will be subleading in our uncertainty.

---

$^{17}$We actually carry this out with the 90% confidence-level NLO MSTW 2008 PDF sets, and then rescale the result, formally a $2\sigma$ variation, by 1.645 to obtain a formally $1\sigma$ variation. This is almost the same as using the 68%-level confidence sets, but because of non-Gaussian tails gives a slightly more conservative estimate of uncertainties.

$^{18}$For example:

$$\Delta_+(R_{1a}) = \sqrt{\sum_i \left( \max \left[ 0, \frac{d\sigma(Z\gamma,S_i^+)}{d\sigma(\gamma\gamma,S_i^+)} - \frac{d\sigma(Z\gamma,S_0)}{d\sigma(\gamma\gamma,S_0)}, \frac{d\sigma(Z\gamma,S_i^-)}{d\sigma(\gamma\gamma,S_i^-)} - \frac{d\sigma(Z\gamma,S_0)}{d\sigma(\gamma\gamma,S_0)} \right] \right)^2}.$$
ties since \( gg \rightarrow V_1^0 V_2^0 \) is itself sufficiently small.

Now we turn to uncertainties in our NLO calculation from renormalization and factorization scales \( \mu_R, \mu_F \). Typically the cancellation of correlated scale variations in ratios of various processes should be viewed as accidental, since the actual structure of higher-order corrections in differing processes is uncorrelated. We wish to argue that this is not the case here. The renormalization scale is sensitive to the ultraviolet region of higher-order corrections, where EW symmetry is restored (up to longitudinal polarizations, which first appear at NNLO in \( gg \rightarrow \phi^3 \phi^3 \)), and where we expect higher-order corrections in general to take a nearly identical form for all \( V_1^0 V_2^0 \) processes. Meanwhile, factorization scale sensitivity primarily comes from divergences associated with emissions off the initial state. While this is not directly affected by the restoration of EW symmetry, it is sensitive to the color structure of the processes order-by-order in the perturbative expansion of QCD, which is also identical for the three \( V_1^0 V_2^0 \) processes. For these reasons the cancellation of scale dependence we observe in our ratios is physical, since the scale choices really are probing correlated higher-order effects.

As shown in figure 2.13, scale-dependence is reduced from several percent in the cross sections to 1–2% in the ratios, where the cancellation is significant for all three ratios and works best at high energy. Here we have varied the scales \( (\mu_R, \mu_F) \) independently from \( \frac{1}{2} m_{VV} \) to \( 2 m_{VV} \) and plotted the envelope of the relative variation in each quantity. However, in figure 2.13 we have held the scales in the \( gg \rightarrow V_1^0 V_2^0 \)
Figure 2.13: The relative uncertainty band on the $V_1^0 V_2^0$ cross sections (left) and $R_1$ ratios (right) found by varying the renormalization and factorization scales $\mu_R, \mu_F$ up and down by a factor of 2. Here the scales appearing in the $gg \rightarrow V_1^0 V_2^0$ process are not varied; see figure 2.14 below.

processes fixed. The calculation to NLO of $q\bar{q} \rightarrow V_1^0 V_2^0$ begins at $O(\alpha_s^0)$, while the calculation of $gg \rightarrow V_1^0 V_2^0$ begins at $O(\alpha_s^2)$. To the order we are working there are no terms in the former calculation which are at the same order as terms in the latter, and thus there is no sense in which the perturbative expansion of the one can affect that of the other. Correspondingly there is no sense in which these two calculations must or should be evaluated with the same value of $\mu_R$, and so their $\mu_R$ dependence must be computed separately. While in principle there could be correlation in the $\mu_F$–dependence through the pdfs, it turns out that $gg \rightarrow V_1^0 V_2^0$ depends much more strongly on $\mu_R$, and so any such correlation is unimportant.

Based on this reasoning, we have also computed the effects of scale variations on the $gg \rightarrow V_1^0 V_2^0$ component of the cross sections, holding all other components fixed. Lacking the $O(\alpha_s^3)$ differential cross sections for $gg \rightarrow Z\gamma$ and $gg \rightarrow ZZ$, we again rely
Figure 2.14: The relative error band on the $V_1^0V_2^0$ cross sections (left) and $R_1$ ratios (right) found by varying $\mu_R, \mu_F$ up and down by a factor of 2. Here only the scales appearing in $gg \rightarrow V_1^0V_2^0$ are varied.

on another incarnation of eq. (2.4.5):

$$d\sigma(3)(gg \rightarrow Z\gamma, \{\mu_1\}) \approx d\sigma(3)(gg \rightarrow \gamma\gamma, \{\mu_1\}) \frac{d\sigma(2)(gg \rightarrow Z\gamma, \{\mu_2\})}{d\sigma(2)(gg \rightarrow \gamma\gamma, \{\mu_2\})},$$

where $\{\mu_i\}$ stands for a choice of $\mu_R$ and $\mu_F$. The resulting uncertainties due to scale variation of the $gg \rightarrow V_1^0V_2^0$ processes are shown in figure 2.14; these are consistent with our estimate from section 2.4.3. Although small for each individual channel compared to the scale variation in the left-hand plot of figure 2.13, cancellations are not as significant as for the NLO scale variations. Consequently the two classes of scale variation turn out to be quite similar in size and shape for the $R_1$ observables, as can be seen in the right-hand plots of figure 2.13 and figure 2.14.

Overall, we can see that while the PDF and scale uncertainties form a significant portion of the theoretical error budget for individual cross sections, these uncertainties are substantially reduced in ratios (in particular in $R_{1\alpha}$) and become subleading.
This presumably reflects true symmetry-related cancellations in the many NNLO corrections that are common to the three neutral diboson processes.

2.4.5 EW corrections

2.4.5.1 Sudakov enhancements

For the level of precision we pursue, higher-order EW corrections to our ratio observables are important. Complete calculations of NLO EW effects for $\gamma\gamma$, $Z\gamma$, and $ZZ$ exist, though public code is not yet at our disposal and the results have been presented with different cuts from our own. As an approximation of the EW corrections, and to estimate the magnitude of their uncertainties, we employ a leading-log calculation in the threshold limit. Comparison of our results below with the full NLO calculations of refs. [87, 158] reassures us that our estimates are reasonable.

Because of various sources of $SU(2) \times U(1)$ breaking, large EW logarithms do not entirely cancel even in fairly inclusive observables such as $d\sigma/d\bar{m}_T$. At very large $\bar{m}_T$, ignoring finite NLO EW corrections and resumming the leading Sudakov logarithms, of the form $\alpha^n \log^{2n} ([\bar{m}_T/m_{W,Z}]^2)$, is justified and should give a good approximation of the dominant effects.

An estimate of the Sudakov logarithm-enhanced corrections can be obtained from a calculation at threshold, where all the energy of the initial state goes into production of the electroweak states. The threshold limit corresponds to a strict veto on the
real emission of EW bosons, so at high $\bar{m}_T$ it overestimates the true EW correction. Since we do not have such a strict veto in our observables, the large virtual corrections above are reduced by our partial inclusion of the real radiation of gauge bosons. For instance, soft $W$ and $Z$ bosons are partially included: a soft $Z$ or $W$ that decays hadronically typically produces soft daughters at wide angles to the hard boson, and thus its daughter jets will neither fail our jet cuts nor ruin isolation of the boson or its daughter leptons. Leptonic decays of the soft bosons are potentially more subtle, depending on how the extra leptons are treated experimentally. Our less extreme veto of soft-collinear bosons should lead to some reduction of the soft-collinear corrections.

Conversely, finite NLO corrections that we ignore in our estimates should increase the size of the EW correction. For moderate values of $\bar{m}_T$, this effect may partially compensate the above-mentioned reduction. Our estimates below are therefore rough guides, and the issue deserves further study.

This threshold regime was studied in the context of boson + jet production [62]. It was found that the EW corrections reduce the photon + jet cross section by $\Delta \sigma_{EW} = -6\%_{-2\%}^{+3\%} (-11\%_{-2\%}^{+3\%})$ at $p_T^\gamma = 500$ (1000) GeV, while reduction of the $Z +$ jet cross section is roughly double this, $\Delta \sigma_{EW} = -13_{-1\%}^{+4\%} (-22_{-1\%}^{+4\%})$. The difference between $Z$ and $\gamma$ arises mainly from loops involving $W$ bosons.

As these effects are primarily associated with the phase space collinear to the hard

---

\footnotesize
\textsuperscript{19}We thank T. Becher for extensive discussions and Xavier Garcia i Tormo for providing detailed results of their calculation.
boson, we anticipate the effect on $\gamma\gamma$ to be roughly the square of the effect on $\gamma + \text{jet}$, leading to a 12–21% reduction in $\sigma(\gamma\gamma)$ for $500 \text{ GeV} < p_T^\gamma < 1000 \text{ GeV}$. Similarly, we expect reductions in $\sigma(Z\gamma)$ [$\sigma(ZZ)$] by 18–31% [24–39%]. But these effects partly cancel in the $R_1$ ratios, reducing $R_{1a}$ ($R_{1b}$) [$R_{1c}$] by just 7–12% (14–23%) [7–12%] in this $p_T$ range. At high enough $p_T$, EW effects become the leading correction to our ratios, dominating over QCD effects.

Importantly, the uncertainties on these EW corrections are not large and are further reduced in our ratios. There are several scale choices which appear in the calculation of ref. [62], but the scale dependence of photons and Zs is correlated, as can be seen in figure 3 of that paper. This correlation reduces the uncertainty in the EW corrections to our ratios. We estimate that the NLO EW uncertainty from scale choices that propagates into our ratios $R_{1a}$ ($R_{1b}$) [$R_{1c}$] is no more than $\pm 2\%$ [$\pm 3\%$ [$\pm 2\%$] for $p_T \sim 500$–1000 GeV. These uncertainties are comparable in size to the uncertainties from PDFs and unknown QCD corrections.

At lower values of $\hat{m}_T$, the finite NLO EW corrections become important, but our resummation approximation still serves as a rough guide to their magnitudes. For $\sigma(\gamma\gamma)$ and $\sigma(ZZ)$, ref. [87] has calculated these corrections as functions of $p_T$. The EW correction is dominated by a logarithmically growing component over much of the $p_T$ range relevant for our ratios, suggesting that our approximation remains applicable in this region. Moreover, comparison of ref. [87] to an earlier calculation of the $\alpha \log^2([p_T/m_{W,Z}]^2)$ term alone [23], corresponding to truncation of the resummed
calculation to first nontrivial order, found agreement at the several percent level. For
similar cuts to ours, ref. [87] claims reductions in $\sigma(\gamma\gamma)$ [$\sigma(ZZ)$] by 13–21% [39–60%]
over the range $p_T \sim 500$–1000 GeV. These reductions are somewhat larger than the 
one we obtained, and resummation is undoubtedly an important part of the discrep-
ancy. At somewhat lower $p_T$, only $\sigma(ZZ)$ shows a clear subleading $p_T$-independent 
correction, which will certainly shift the EW corrections to $R_{1b}, R_{1c}$ away from our 
leading-log predictions.

NLO EW results for $\sigma(Z\gamma)$ are given in ref. [158], but only with a fixed and low 
cut on $p_{T,Z}$. This makes comparison with our estimates impossible, because large 
logarithms of $p_{T,\gamma}/p_{T,Z}^{\text{cut}}$ arise and are indistinguishable from inclusive EW Sudakov 
logarithms. Still, we have no reason to suspect that the behavior of the finite EW cor-
rections should be qualitatively different from those of $\gamma\gamma$ and $ZZ$.

Most importantly for our purposes, when finite pieces numerically dominate the 
NLO EW correction, its uncertainty arises mainly from scale variation in the EW cou-
plings. Our earlier estimate of the uncertainty using ref. [62] is therefore an overesti-
mate at small $\bar{m}_T$.

We have summarized these statements in figure 2.1 of section 2.2 by indicating the 
expected fractional shifts in the ratios due to the source of EW corrections derived in 
ref. [62], along with an estimate of their uncertainties. This shows that these EW ef-
effects might be observable in our ratios in the highest bins, where they dominate QCD 
effects. Furthermore, EW effects are under sufficient control that there will still be
substantial sensitivity to other, non-SM contributions at high $\tilde{m}_T$.

2.4.5.2 Proper choice of EW scales for on-shell external photons

Another EW issue concerns the correct choice of electromagnetic coupling corresponding to emission of a photon.\textsuperscript{20} In the literature one finds preference for evaluating $\alpha(\mu_{\text{QED}})$ both at $\mu_{\text{QED}} = 0$ and at $\mu_{\text{QED}} = \min(m_Z, \sqrt{s})$ (or some fraction thereof). Since the QED coupling runs by 7% from 0 to $m_Z$, this difference affects $R_{1a}$ and $R_{1c}$ by 7% and $R_{1b}$ by 14%.

Typical QCD calculations may seem to suggest using $\mu_{\text{QED}} \sim \sqrt{s}$. But in contrast to a quark or gluon, we can experimentally require that a photon is on-shell and does not shower, i.e., does not form an electromagnetic jet of leptons and hadrons with a finite mass. For abelian gauge bosons, the leading effect of requiring an \textit{on-shell} photon, rather than a photon that could be off-shell by as much as $q^2 \sim \hat{s}$, is given by running the coupling down from $\mu_{\text{QED}} = \sqrt{s}$ to $\mu_{\text{QED}} = 0$. (Importantly this is not true for nonabelian gauge bosons.) This choice removes photons that, for instance, split to a $\mu^+\mu^-$ pair or mix with the $\rho$. We find this argument reliable in a pure color-singlet situation, such as Higgs decay to two photons.

Subtleties could arise, however, in a colored environment: soft ISR gluons are present in $pp$ collisions and can be radiated into the photon isolation cone. On the one hand,

\textsuperscript{20}We thank Z. Bern for pointing out the issue, and for conversations.
we still want to forbid $\gamma^* \rightarrow \mu^+\mu^-$ since this would be experimentally rejected; this
tends to suggest $\mu_{\text{QED}} < 2m_\mu$. On the other hand, we should include photons with
nearby soft gluons that lie below the isolation cut $p_{T,\text{min}}^{\text{had}}$, which could suggest$^{21}$ $\mu_{\text{QED}} \sim p_{T,\text{min}}^{\text{had}}$.

Faced with a lack of consensus, we have chosen not to directly address this issue
in this paper. Instead we use MCFM 6.8 “out of the box”, for which $\mu_{\text{QED}} = m_Z$
throughout. In figure 2.1 of section 2.2, we have indicated the potential shift from
switching to $\mu_{\text{QED}} = 0$ as an overall 7% or 14% error band that is essentially flat and
fully correlated across all bins. (Even if this dispute were not resolved theoretically,
the measurement of the average ratio of the lowest bins would largely fix the value of
$\mu_{\text{QED}}$.) In no sense should this be thought of as a Gaussian error band, since no prob-
ability extends beyond the band. For now readers may adjust our results according to
their individual opinions, but clearly it is important that consensus on the matter be
reached in the near future.

$^{21}$Suggested to us by T. Becher following ref. [62]. A related suggestion was made
by M. Schwartz.
2.5 Additional practical considerations

2.5.1 Photon isolation

In section 2.4 we used the smooth-cone photon-isolation method of Frixione, eq. (2.4.1), but this is experimentally impractical. More traditional is hard-cone isolation, simply requiring that the energy in a cone of size $R_h$ around the photon be less than $\epsilon_h p_T^\gamma$.

But if $\epsilon_h$ is small, a hard cone produces large logarithms due to the incomplete cancellation of virtual and soft gluon effects. Meanwhile if $\epsilon_h$ is not small, the hard cone introduces large sensitivity to the fragmentation function $D_{q\to\gamma}(z)$ at $z \to 1$, which is dangerous to a precision calculation since $D_{q\to\gamma}(z \to 1)$ has substantial associated uncertainties. The Frixione algorithm avoids these issues by removing the divergent regions of phase space that require the introduction of a fragmentation function in the first place. The isolation parameters can then be set so that no large perturbatively calculable logarithms appear. However, the smooth cone cannot be implemented experimentally since it requires the energy in a small cone around the photon to go literally to zero as that cone decreases in size. This difficulty may be evaded by using a discretized or “staircase” version of the smooth cone [88, 212]. Although sensitivity to the photon fragmentation function is thereby reintroduced, this sensitivity can be maintained small while keeping the associated logarithms of manageable size, so as to not call the accuracy of the fixed-order calculation into question.

Our staircase isolation approximates the smooth cone of eq. (2.4.1), which has pa-
<table>
<thead>
<tr>
<th>$R$</th>
<th>$\epsilon_h$</th>
<th>$E_{\text{min}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>5 GeV</td>
</tr>
<tr>
<td>0.2</td>
<td>0.07</td>
<td>10 GeV</td>
</tr>
<tr>
<td>0.3</td>
<td>0.20</td>
<td>23 GeV</td>
</tr>
<tr>
<td>0.4</td>
<td>0.38</td>
<td>40 GeV</td>
</tr>
</tbody>
</table>

Table 2.4: Four concentric hard cones used to approximate smooth-cone isolation. $R$ is the cone angle, $\epsilon_h$ is the energy fraction, and $E_{\text{min}}$ is a threshold below which we do not reject events, regardless of hadronic energy fraction in the cone. Note that the value $\epsilon_h^{(1)}$ is so small that, in our kinematic regime, isolation in the innermost cone is always controlled by the energy cutoff $E_{\text{min}}$.

We choose four nested cones ($n = 1, 2, 3, 4$) with radii $R_h^{(n)} = 0.1 \times n$, and approximate the function $I(R; \epsilon, \delta)$ of eq. (2.4.2) by a piecewise constant function

$$\hat{I}(R; \epsilon, \delta) = \epsilon \left[ 1 - \cos \left( \frac{1}{2} (R_h^{(n)} + R_h^{(n-1)}) \right) \right] \equiv \epsilon_h^{(n)}, \quad \text{for} \quad R_h^{(n-1)} < R < R_h^{(n)},$$

where we define $R_h^{(0)} = 0$. The constants $\epsilon_h^{(n)}$ are shown in table 2.4; the functions $I$ and $\hat{I}$ are plotted at left in figure 2.15. Then our staircase isolation criterion requires

$$\sum_{h \in R^{(n)}} p_T^h < \max \left\{ \epsilon_h^{(n)} p_T, E_{\text{min}}^{(n)} \right\},$$

where the energies $E_{\text{min}}^{(n)}$, given in table 2.4, are chosen so that they lie at or above the expected level of pile-up (up to an average of 60 pp collisions per crossing) over Run 2 and 3 of the LHC. Since event-by-event pile-up subtraction techniques will remove a significant fraction of the energy deposited in the isolation cone, this choice will assure that our technique will not suffer from large efficiency losses due to pile-up.
Figure 2.15: Comparing staircase isolation to the Frixione algorithm. (Left) The smooth curve is $I(R; \epsilon, \delta)$ of eq. (2.4.2), while the piecewise-constant curve is $\tilde{I}(R; \epsilon, \delta)$ of eq. (2.5.1). (Right) The effect, on $\sigma(\gamma\gamma)$ at NLO, of changing the isolation procedure. Here, $\sigma$(smooth) corresponds to pure Frixione isolation, eq. (2.4.1), while $\sigma$(stair) is computed using eq. (2.5.2). At high energies, staircase isolation is indistinguishable from the Frixione algorithm; even at low energies, the difference is slight.

At right in figure 2.15, we compare our staircase isolation with the Frixione algorithm, by computing $\sigma(\gamma\gamma)$ with each isolation method and taking the relative difference of the results. The two methods differ by at most 4% [2%] in $\sigma(\gamma\gamma)$ [$\sigma(Z\gamma)$], and the difference decreases with energy. Staircase isolation thus shifts the central value of $R_{1a}$ ($R_{1b}$) [$R_{1c}$] up by at most 2% (4%) [2%] from the values computed in section 2.4 with smooth-cone isolation.

Now, having seen that the two photon-isolation procedures are not substantially different for our ratios, let us discuss the uncertainties associated with the staircase method. One source of uncertainties stems from the experimental extraction of the fragmentation function. We use the leading-order $q \rightarrow \gamma$ fragmentation function,
since our NLO calculations involve working only to leading order in $q \to q\gamma$ splitting. The photon fragmentation function for a quark parent has been measured most precisely at ALEPH [94], in $Z \to \gamma$ + hadrons, in which the final state is dominated by $Z \to q\bar{q}\gamma$ and the fragmentation function contributes to the region where a quark or antiquark becomes collinear with the photon. The function extracted at leading order by ALEPH, based on a QCD analysis proposed in ref. [201], is

$$D^{\text{LO}}_{\gamma\to q}(z, \mu_0) = \frac{\alpha_{\text{em}}^2}{2\pi} \left( \frac{P^{(0)}_{\gamma\to q}(\log \frac{\mu_F^2}{\mu_0^2(1-z)^2} + C)\right), \quad (2.5.3)$$

$$\mu_0 = 0.22^{+1.3}_{-0.19} \text{ GeV}, \quad (2.5.4)$$

$$C = -12.1 \pm 4.3, \quad (2.5.5)$$

where $P^{(0)}_{\gamma\to q}$ is the tree-level perturbative splitting function. Uncertainties on the two parameters appear large at first glance, but the parameters are highly correlated. ALEPH suggested that one should take the relation

$$C = -1 - \log \left( \frac{s}{2\mu_0^2} \right) \bigg|_{s=m_Z^2}, \quad (2.5.6)$$

and found

$$\mu_0 = 0.14^{+0.21}_{-0.08} - 0.04 \text{ GeV}, \quad C = -13.26. \quad (2.5.7)$$

This uncertainty in $\mu_0$ propagates into a minute (per-mil) uncertainty in our ratios.

But since the correlation in eq. (2.5.6) is not assigned an uncertainty, this approach is slightly over-optimistic. On the other hand we can obtain an overly-conservative estimate if we ignore the correlation and vary both parameters independently by the
uncertainties listed in eqs. (2.5.4) and (2.5.5). In this case we find uncertainties of about 1% on our ratio $R_{1a}$. As this is surely a considerable over-estimate, we believe that this source of uncertainty is unimportant.

Several other sources might inflate the uncertainties of the isolation contribution if not handled correctly. First is the fact that, working to NLO in $V_1^0V_2^0$ production, we have done only a leading-order calculation for quark-photon splitting (and used the corresponding LO fragmentation function). However, since the sensitivity to the fragmentation function is minimized by the staircase method, we do not think the next order correction will affect our ratios in a material way. At the same time, since none of the currently available fragmentation function fits perform any resummation of the logarithms of $\log^2(1 - z)$ that appear in the perturbative fragmentation contribution, one must be careful to implement isolation in such a way that one does not weight the $z \to 1$ region of phase space too strongly. The staircase isolation that we advocate here does precisely this, in contrast to hard-cone isolation with a small radius.

### 2.5.2 $Z$ decay and lepton isolation

Up to this point we have treated the $Z$ as though it does not decay, and imposed the same cuts on $\gamma$ and $Z$ as shown in table 2.3. But the decay of the $Z$ forces us to impose kinematic and isolation cuts on its daughter leptons and to account for the $Z$ peak’s width in defining what we mean by a $Z$. (We consider non-leptonic decays of
<table>
<thead>
<tr>
<th>Kinematic Cuts</th>
<th>Isolation Cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>m_{\ell\ell} - m_Z</td>
</tr>
<tr>
<td>$p_T^{\ell_1} &gt; 20 \text{ GeV}$</td>
<td>$\Delta R_{\ell^+\ell^-} \geq 0.0$</td>
</tr>
<tr>
<td>$p_T^{\ell_{2,3,4}} &gt; 7 \text{ GeV}$</td>
<td>$p_T(j) &lt; 0.2 \times p_T(\ell)$</td>
</tr>
<tr>
<td>$</td>
<td>\eta(\ell)</td>
</tr>
</tbody>
</table>

Table 2.5: Kinematic and isolation cuts imposed on daughter leptons. The leptons are $p_T$-ordered such that $p_T^{\ell_1} > p_T^{\ell_2} > p_T^{\ell_3} > p_T^{\ell_4}$.

the $Z$ briefly in section 2.6.2.) This has a significant though highly predictable effect on the measurements.

To a good approximation we find that the effects of these three experimental realities factorize, meaning that the overall acceptance $\zeta$ of these three effects can be written as the product of separate acceptance factors:

$$
\zeta = \zeta_{\gamma,\Delta} \times \zeta_{\text{kin}} \times \zeta_{\text{iso}}, \quad (2.5.8)
$$

where the $\zeta_i$ are defined as a relative change to the cross section due to a particular effect: $\zeta_{\gamma,\Delta}$ is the acceptance after requiring the dilepton mass be within $\Delta$ of the $Z$ pole, $\zeta_{\text{kin}}$ is the acceptance of our lepton kinematic cuts, and $\zeta_{\text{iso}}$ is the acceptance of our lepton isolation cuts. Let us now discuss each of them in turn.

The $Z$’s finite width and $Z-\gamma^*$ interference require that we define what we mean by a $Z$ boson. We take a $Z$ to be an opposite-sign same-flavor dilepton pair whose mass $m_{\ell\ell}$ falls within $\Delta = 25 \text{ GeV}$ of $m_Z = 91.2 \text{ GeV}$. To quantify effects of this mass
window, we define $\zeta_{\Gamma, \Delta}$ as the ratio of a finite-width cross section with mass window $\Delta$ divided by the (fictitious) zero-width cross section that we have used up to now. Note $\zeta_{\Gamma, \Delta}$ can exceed 1 if the window $\Delta$ is taken sufficiently wide.

Also at this stage, to remove a divergence in the $pp \rightarrow Z\gamma$ cross section, we apply an isolation cut between leptons and photons by requiring that $\Delta R_{\ell\gamma} > 0.2$ for any lepton-photon pair in the event. We find that if we change $\Delta R_{\gamma\ell} > 0.2$ to $\Delta R_{\gamma\ell} > 0.4$, the cross section changes by less than 0.5%. This is unsurprising since the kinematic cuts of table 2.3 force the $Z$ and $\gamma$ to be well-separated. The small effect of this isolation cut is included in $\zeta_{\Gamma, \Delta}$.

We next impose realistic kinematic cuts on individual leptons, as shown in the kinematic cuts section of table 2.5. In order to quantify the effects of kinematic cuts on individual leptons, we define $\zeta_{\text{kin}}$ as the ratio of cross sections with and without these kinematic cuts.

Although the $Z$ is not itself observed, we retain the cut on $Z$s shown in table 2.3; that is, we reject $Z$s with $|\eta| > 1.5$ even if the leptons pass the cuts in table 2.5. This choice is somewhat arbitrary and may not be necessary, but making different $\eta$ cuts on $Z$s and $\gamma$s might inflate PDF uncertainties, which otherwise have substantial cancellations.

Finally, we impose isolation cuts between the daughter leptons of a $Z$ and all the jets in the final state. Specifically, we require that $p_T(j) < 0.2 \times p_T(\ell)$ for any jet-lepton pair with $\Delta R_{\ell j} < 0.4$. The effect of these isolation cuts is described by $\zeta_{\text{iso}}$. We
Figure 2.16: Effects of finite $Z$ width and lepton cuts on $Z\gamma$ (left) and $ZZ$ cross sections (right). Each plot is normalized by the respective cross section with no lepton cuts and $\Gamma_Z = 0$. The red circles show the effect of the $Z$ width. The green triangles combine the width with the kinematic cuts of table 2.5 and photon-lepton isolation cuts. The blue squares combine these with the other lepton isolation cuts of table 2.5. See text for more details and notation.

define $\zeta_{\text{iso}}$ as the ratio of cross sections with and without the jet-lepton isolation cuts.

Our $Z$ bosons are often boosted. To avoid unnecessary acceptance losses, we do not require any lepton to be isolated from another lepton of the same flavor and opposite sign. We do not expect this to cause a large $Z$ fake rate at the relevant $\hat{m}_T$.

As shown in figure 2.16, these cuts lower the $Z$ acceptance; for $Z\gamma$ ($ZZ$) events, acceptance drops to 94% (85%) for the lowest values of $\hat{m}_T$, rising toward 100% (90%) at higher values. Losses are small at high $\hat{m}_T$ because the leptons have large $p_T$ and have similar $\eta$ to the parent $Z$. Losses would be much greater (78% and 58% acceptance for $Z\gamma$ and $ZZ$) if all four leptons were required to have $p_T > 20$ GeV.

These effects thus change our ratios, reducing cross sections by 5–8% for each $Z$. However these effects are calculable and do not increase uncertainties. There should
be no problem to include them in the theoretical predictions or unfold them from the experimental measurements.

2.6 Discussion and summary

2.6.1 Uncertainty budget

<table>
<thead>
<tr>
<th>Effect</th>
<th>$R_{1a}$ (Zγ/γγ)</th>
<th>$R_{1b}$ (ZZ/γγ)</th>
<th>$R_{1c}$ (ZZ/Zγ)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$qq \rightarrow VVqq$</td>
<td>2–3%</td>
<td>3–3.5%</td>
<td>1.5–2.5%</td>
<td>section 2.4.2</td>
</tr>
<tr>
<td>$\mu_R, \mu_F$ (gg)</td>
<td>0.5–1%</td>
<td>1%</td>
<td>1–2%</td>
<td>section 2.4.4</td>
</tr>
<tr>
<td>$\mu_R, \mu_F$ (NLO)</td>
<td>0.5–1%</td>
<td>1.5–2.5%</td>
<td>1–1.5%</td>
<td>section 2.4.4</td>
</tr>
<tr>
<td>PDF</td>
<td>0.5%</td>
<td>1–1.5%</td>
<td>0.5–1%</td>
<td>section 2.4.4</td>
</tr>
<tr>
<td>EW (LL)</td>
<td>+2%</td>
<td>+3%</td>
<td>+2%</td>
<td>section 2.4.5.1</td>
</tr>
<tr>
<td>$\alpha_{QED}$</td>
<td>7%</td>
<td>14%</td>
<td>7%</td>
<td>section 2.4.5.2</td>
</tr>
</tbody>
</table>

Table 2.6: Summary of overall uncertainty budget. The first three entries are not independent sources of uncertainty, and combining them assuming no correlation provides a conservative estimate.

In section 2.2, we presented our claim that the three $R_1$ ratios (whose central values are related but which have different cancellations among their uncertainties) are under exceptional theoretical control. Here we present a detailed breakdown of what we include in our estimate of known theory uncertainties, as shown in table 2.6, and justify our confidence in the small size of further higher-order effects. We now review
the table line by line.

The first three lines of table 2.6 are not truly independent, as they are all striving to capture aspects of the uncertainty associated with higher-order corrections to our calculations of the ratios. Our goal in isolating them was to try to identify any particularly large effects, ones that would not show up in overall NLO scale variations, that we have not already included and would not cancel in our ratios. Although the separation we have made is both scheme and scale dependent and thus unphysical, our methods are probably sufficient to estimate the rough magnitude of the higher-order corrections that we did not include. We have also been quite conservative in our estimates and in how we combined uncertainties. Once NNLO calculations of all diboson processes become publicly accessible, the uncertainties from all sources should be subsumed in the scale variation of the analogous NNLO calculations, with the exception of the $gg$ initial state, which only first appears at NNLO. For this last part, two-loop results for $\gamma\gamma$ and $ZZ$ already exist [80, 106], as do most components of the $Z\gamma$ calculation [195], allowing for a more robust characterization of the associated uncertainties than the estimates we have performed here.

As we noted in section 2.4.2 above, many NNLO corrections are expected to cancel in the $R_1$ ratios. Valence quark scattering $qq \rightarrow V_1^0V_2^0qq$, which has terms that are not proportional to the LO cross sections, gives one of the largest non-canceling terms that we cannot currently compute. Our method for obtaining these estimates was described in section 2.4.3.
We obtained estimates of the $O(\alpha^4)$-uncertainty in $gg \to \gamma\gamma$ production by varying the scales $\mu_R, \mu_F$ in $gg \to \gamma\gamma$, computed to $O(\alpha^3)$, up and down by a factor of two. Because of $SU(2) \times U(1)$ relations, we assumed that nearly the same relative uncertainty applies to $gg \to Z\gamma$ and $gg \to ZZ$. See section 2.4.4 for more details.

Although it is possible that there are other large non-cancelling NNLO effects, we have not been able to identify them. In particular, although collinear effects make a large contribution at NLO, their contribution at NNLO appears to be much smaller. Moreover, there are no other new channels or new regions of phase space that open up at this order. Consequently we naively expect other NNLO shifts to the ratios to largely cancel. We estimated these effects in section 2.4.4 by seeing how varying the renormalization and factorization scales for the strictly NLO calculations independently affect the ratios; the $\sim 5\%$ corrections in each channel have very substantial correlations, and largely cancel in the ratios.

PDF uncertainties are extracted from the calculations that led to figure 2.12. We find that they are very small for $R_{1a}$, and even for the other ratios are significantly smaller in percentage terms than for the individual diboson processes.

Now we turn to the EW uncertainties. The leading-log EW uncertainties, dominated by the choice of matching scales, were extracted from the threshold resummation calculation of ref. [62] as described in section 2.4.5. We also account for the differing views of how to set the scale for $\alpha(\mu_{\text{QED}})$ by varying $\mu_{\text{QED}}$ between 0 and $m_Z$. Note this is a window and not in any sense a 1$\sigma$ Gaussian uncertainty.
One item for which we do not have an error estimation is photon isolation. An essential part of our proposal involves the use of staircase isolation, an experimentally practical approximation to the Frixone smooth-cone method, discussed in section 2.5.1. The use of a hard cone for isolation would introduce a substantial shift to our result and significant sensitivity to $q \rightarrow \gamma$ fragmentation. Staircase isolation minimizes these effects. We saw in section 2.5.1 that the difference between smooth and staircase cone, most important at low $\bar{m}_T$, is at most 2% for $R_{1a}$ and $R_{1c}$, and double this for $R_{1b}$. The effect of experimental uncertainties on the fragmentation function, which we estimated by varying the parameters in ALEPH’s fit, appears to be negligible.

2.6.2 Final comments

2.6.2.1 General reflections on our methods

We have proposed a wide variety of ratios using LO reasoning about the $SU(2) \times U(1)$ structure of the SM. Interestingly, the structure of $SU(2) \times U(1)$ and the radiation zero in the amplitude $a_3$ means that the naive guess for custodial-$SU(2)$ relationships among $W$ and $Z$ do not hold. The only interesting relation between $ZZ$, $W^\pm Z$ and $W^- W^+$ production is an imperfect (and somewhat impractical) relation between

\[ \frac{R_{1a}}{R_{1c}} = \frac{\sigma(W^- W^+)}{\sigma(ZZ)} \frac{1}{\sigma(W^+ W^-)} \]

\[ \frac{R_{1b}}{R_{1c}} = \frac{\sigma(W^- W^+)}{\sigma(ZZ)} \frac{1}{\sigma(W^+ W^-)} \]

22Note that uncertainties in the fragmentation function might be reducible. The ALEPH measurement could be repeated at the LHC, using $W$ decays arising in $t\bar{t}$ events. Selecting events with a lepton, $E_T$, two $b$ tags and a loose photon, one could then reconstruct the tops and extract the probability that $W \rightarrow \gamma +$ hadrons.
\[ W^- W^+ \text{ and } ZZ, \text{ which follows only because } |a_3|^2 \text{ is subdominant in } W^- W^+ \text{ production. This could be generalized to a relation between } W^- W^+, W^\pm Z \text{ and } ZZ, \text{ but the relation is complicated as well as impractical. We also saw no interesting relation between } W^\pm \gamma \text{ and } Z\gamma. \text{ That said, charge ratios of } W^+ Z \text{ to } W^- Z, \text{ and of } W^+ \gamma \text{ to } W^- \gamma, \text{ are important tools. Although we focused on diboson ratios at high } \bar{m}_T, \text{ the nice properties of these charge ratios do not require } SU(2) \times U(1), \text{ and would remain interesting even down to low } \bar{m}_T. \]

An issue that we have not addressed is the experimental systematic uncertainty from fake photons. We have seen in the Higgs boson search that the experiments have fairly large contributions to their \( \gamma \gamma \) searches from \( \gamma + \text{jet} \). Although these decrease at moderate \( \bar{m}_T \), partially cancel in our ratios, and are typically smaller for photons in the barrel of ATLAS and CMS, they are by no means negligible, as can be seen in ref. [232]. We have implicitly assumed that the systematic uncertainties from these fakes will be under very good control for \( p_T^\gamma \geq 150 \text{ GeV} \) by the time 300 fb\(^{-1}\) has been accumulated. If this is not true it could make the \( R_1 \) ratios, and others we have proposed, somewhat less useful.

We have limited our detailed study to NLO QCD effects, for practical reasons. Many NNLO QCD and higher order EW corrections have been performed already, so it should soon be possible to improve upon our results and, most importantly, check our uncertainty estimates. At NNLO, with two jets accompanying the two bosons, one would encounter many new issues, including vector boson scattering and potential...
sensitivity to new phenomena therein, as well as $SU(2)$-quintet amplitudes, including same-sign $WW$ production. However, few of these issues may be essential in the ratios we propose, since the low rates for diboson production mean that theoretical predictions more precise than a few percent may often not be needed at the LHC.

Our results for the $R_1$ ratios involved many arbitrary choices including specific kinematic cuts, isolation requirements, binning, etc. Although we have carefully considered these choices, we have not in any sense optimized them, and further consideration, both theoretical and experimental, should be given to them.

Finally, in our results we have imposed an isolation cone around photons but not around $Z$s. This appears to be a sub-percent effect (for each $Z$) with the isolation criteria that we selected. However, at a higher energy collider this must be revisited, since at sufficiently high energy the $Z$ and photon will have to be treated on equal footing and a photon-like isolation on the reconstructed $Z$ will have to be applied; otherwise large EW logarithms will afflict our ratios.

### 2.6.2.2 Other $Z$ decays

The $R_1$ ratios, especially the two involving $ZZ$, suffer from low statistics, due to the small $Z \rightarrow \ell^+\ell^-$ branching fraction. One might wonder whether one can gain by looking at $Z\gamma$ events in which the $Z$ decays to neutrinos, and especially at $ZZ$ events by looking for $\ell^+\ell^-$ plus missing transverse momentum ($E_T$). We have not explored this, but in the ATLAS measurement of the $ZZ$ production cross section [13], such
signal events are incorporated.

An obvious downside to this approach would be an inability to put the same $\eta$ cuts on $Z$s and $\gamma$s. Since these $SU(2)$-singlet processes are generated in the $t$ and $u$ channel, they are particularly sensitive to the $\eta$ cut, so having different cuts for $\gamma$ and $Z$ could potentially cause large NLO corrections due to imperfect cancellations. Also, the excellent cancellation of PDF uncertainties in $R_{1a}$ could potentially fail. There would also be backgrounds from $W\gamma$ and $WZ$ events where the $W$ decays to a hadronic tau or a soft lepton, and is mistaken for an invisibly decaying $Z$. The ideal balance between smaller statistical uncertainties and larger theoretical uncertainties will be time-dependent, and requires study by the experimental LHC groups at the time of the measurement.

Nevertheless, there might be a practical strategy using $Z \to \nu\bar{\nu}$ events. One could measure $\gamma\gamma$ differentially, and use the $R_{1a}$ and $R_{1b}$ ratios to predict, in the central region, the $Z\gamma$ and $ZZ$ distributions. Then, assuming the SM, the full $Z\gamma$ and $ZZ$ distributions extending to larger $\eta$ could be predicted with lower uncertainties, and this prediction could then be checked against events with $E_T$ and a single $\gamma$ or leptonic $Z$. $^{23}$

The option of using hadronic decays of the $Z$ seems daunting. The backgrounds from $Z$- or $\gamma$-plus-jet events, where a QCD jet fakes a boosted $W$ or $Z$, are not small,

$^{23}$A similar method is presumably needed for any ratios involving $W^+W^-$ production, where both $\eta$ and especially $\vec{m}_T$ are somewhat uncertain in each event.
and will leak into the diboson measurement. Moreover, mass- and charge resolution on hadronically decaying vector bosons is poor, so one cannot distinguish $Z\gamma$ from $W^{\pm}\gamma$, processes with completely different differential distributions even at LO.

A further tool that we have not explored is the use of $Z$ polarization, potentially of interest due to the parity violation in the SM. BSM physics might alter polarization ratios.

### 2.6.2.3 Applications of the $R_1$ ratios

The $R_1$ ratios should be useful in several ways even within the SM. First, they allow high-precision tests of SM calculations, including Monte Carlo methods. Second, they may serve as a place to explore higher-order EW effects. As we described in section 2.4.5, EW corrections partly cancel in these ratios, but can reach the 10–20\% level, above the level of theoretical uncertainties. The fact that QCD corrections cancel rather completely, especially at high $m_T$, means these ratios may serve as a particularly clean place to examine logarithmically-enhanced EW effects.

In this paper we have not addressed the question of how sensitive these variables would be to BSM phenomena. An obvious potential use of these variables is in searching for BSM interactions of the SM gauge bosons, e.g. through anomalous triple and quartic gauge couplings (aTGCs, aQCSs). In exploring this, one should use an $SU(2)\times U(1)$ invariant classification of the various operators. We leave this for future work. Wide resonances decaying to EW bosons might also alter the ratios without being
observable in some simpler way. Most other phenomena would introduce additional hard jets, which would often be vetoed by our cuts, or large amounts of $E_T$, which would not impact the $R_1$ ratios but could affect ratios involving leptonic $W$s, as well as any measurements that try to use $Z \to \nu\bar{\nu}$ as we outlined above. Strategies for events with additional jets and/or $E_T$ are worth further study, but it is far from clear whether our methods can be suitably generalized to such cases.

At higher collision energy, such as a 30 or 100 TeV $pp$ collider, these observables are probably still useful, but will perhaps be more complicated. On the one hand, finite mass effects will be completely negligible, and the fragmentation contribution for $W$s and $Z$s will become similar to that for photons. EW corrections become quite large and could easily be observed in these ratios, as would any TeV-scale new physics effects. However, other issues, such as the non-negligible rate for a hard lepton to radiate a real EW boson, as well as the general challenges of resolving and identifying gauge bosons at ultra-high boosts, will begin to have a practical impact on precise diboson measurements. A dedicated study of this question is needed.

2.6.2.4 Prospects for the other ratios

What can we expect for the other ratios of eq. (2.2.2) at NLO? The $R_1$ ratios are somewhat special. First, the $\gamma\gamma$, $Z\gamma$ and $ZZ$ processes are fully reconstructible, though at the cost of $Z \to \ell^+\ell^-$ branching fractions. By contrast, $W\gamma$ and $WZ$ events with a single neutrino are only reconstructible up to a two-fold ambiguity, and leptonic $WW$
events cannot be reconstructed event-by-event.

Second, $\gamma\gamma$, $Z\gamma$ and $ZZ$ cross sections are all proportional to the singlet amplitude-squared $|a_1|^2$ at LO, and many of their NLO corrections are identical at high energy. This is not true for the other processes. A particular complication is the fact that the $SU(2)$-triplet amplitude-squared $|a_3|^2$ vanishes at scattering angle $\pi/2$, or in other words at $s = 4m^2_T$. The falling PDFs assure this is a kinematic region of particular importance for production rates at hadron colliders. Differential cross sections for $W^\pm\gamma$, $W^\pm Z$ and $W^- W^+$ are suppressed at LO by this “radiation zero”, but to different degrees, and what remains behind is different in each case. The radiation zero is removed at NLO, and consequently some ratios, particularly certain asymmetries which are quite small at LO, may end up with large NLO corrections and NNLO uncertainties. Indeed it is already well-known that the $K_{NLO/LO}$ factor for $W\gamma$ is much larger than that for $Z\gamma$ [207, 282].

Despite these challenges, there are enough variables in our list that some may evade these concerns. We are optimistic that a few of the remaining variables will be as precisely predictable as the $R_1$ ratios, and we plan to explore this possibility further. In the meantime, we hope that our methods will inspire invention of other precision observables, perhaps more sophisticated and less obvious, for the LHC and for hadron colliders of the future.
Jet substructure beyond the next-to-leading logarithm
3.1 Introduction

The high luminosity proton collisions at the Large Hadron Collider (LHC) enable an unprecedented sensitivity to rare and high scale physics. The cost of such high luminosities is the presence of significant amounts of pile-up radiation present in every event, arising from numerous secondary proton collisions per bunch crossing. Pile-up is truly uncorrelated with the hard scattering and can contaminate any potential measurement. This is particularly important for measurements made on jets, for which pile-up can effect a large systematic bias in observables like the jet mass. In searches for resonances that decay to boosted electroweak objects which have definite masses, pile-up can significantly degrade the ability to separate signal from background. Over the past several years, numerous methods [81, 82, 96, 98, 100, 148, 169, 244, 245, 248, 300] have been developed for grooming jets and events for pile-up mitigation and removal, and are now standard experimental tools at both ATLAS and CMS experiments. Especially in analyses of jets, measurements made at the LHC often involve some form of grooming.

With this motivation, it is imperative to understand these jet grooming techniques from first principles QCD. There have been a few studies of the theoretical aspects of jet groomers [147, 148, 150, 248], with predictions for jet-observable distributions calculated to next-to-leading logarithmic (NLL) accuracy for two widely used jet groomers: the modified mass drop tagger (mMDT) and soft drop. These explicit
analytic studies showed that these jet groomers not only have desired experimental properties, but can also dramatically simplify theoretical calculations as compared to their ungroomed counterparts. Non-global logarithms (NGLs) that arise from correlations between in- and out-of-jet scales have proven to be a significant obstruction to resummation of ungroomed jet observables to NLL accuracy and beyond. In particular, it was demonstrated by explicit calculation in Refs. [147, 148, 248] that mMDT and soft drop groomers eliminate the leading non-global logarithms in jet mass distributions [151]. mMDT and soft drop pave the way for systematically improvable resummed predictions of jet observables.

In this paper, we open the door to systematically improvable jet substructure calculations by presenting an all-orders factorization theorem for the soft-drop [248] groomed observables using soft-collinear effective theory (SCET) [51, 52, 55, 56]. An overview of the method we discuss here and some of our results were presented recently in Ref. [188]. This paper provides a more detailed presentation of those results as well as a derivation of the factorization formula and its remarkable properties.

The soft drop groomer walks through the branching history of a jet, discarding soft branches until a sufficiently hard branching is found. This is enforced by effectively requiring

$$\frac{\min[E_i, E_j]}{E_i + E_j} > z_{\text{cut}} \left( \frac{\theta_{ij}}{R} \right)^{\beta}, \quad (3.1.1)$$

where $E_i$ and $E_j$ are the energies of the particles in that step of the branching, $\theta_{ij}$ is
their relative angle, and $R$ is the radius of the jet. $z_{\text{cut}}$ is a parameter that sets the scale of soft, wide angle emissions in the jet; the typical value is $z_{\text{cut}} = 0.1$. $\beta$ is a parameter that controls the aggressiveness of the groomer: $\beta = \infty$ removes the groomer, $\beta = 0$ coincides with mMDT and is simply an energy cut, and $\beta < 0$ removes all soft and collinear singularities. We will consider $\beta \geq 0$. If Eq. (3.1.1) is not satisfied, the softer of the two branches is removed from the jet, and the grooming procedure continues on the harder branch. When Eq. (3.1.1) is satisfied, the procedure terminates and the groomed jet is returned. For concreteness, on this groomed jet, we measure the two-point energy correlation functions $e_2^{(\alpha)}$ with angular exponent $\alpha > 0$ [47, 221, 253].

In $e^+e^- \to$ dijets events, the factorization formula we derive in this paper for soft-drop groomed left and right hemisphere jets is:

$$\frac{d^2\sigma}{d e_{2,L}^{(\alpha)} d e_{2,R}^{(\alpha)}} = H(Q^2) S_G(z_{\text{cut}}) \left[ S_C(z_{\text{cut}} e_{2,L}^{(\alpha)}) \otimes J(e_{2,L}^{(\alpha)}) \right] \left[ S_C(z_{\text{cut}} e_{2,R}^{(\alpha)}) \otimes J(e_{2,R}^{(\alpha)}) \right]. \quad (3.1.2)$$

This factorization theorem applies when $z_{\text{cut}} \ll 1$ and the left- and right-hemisphere energy correlation functions are asymptotically small: $e_{2,L}^{(\alpha)}, e_{2,R}^{(\alpha)} \ll z_{\text{cut}} \ll 1$. We illustrate the physical configuration corresponding to this factorization theorem in Fig. 3.1. In Eq. (3.1.2), $H(Q^2)$ is the hard function for $e^+e^- \to q\bar{q}$. $S_G(z_{\text{cut}})$ is the global soft function, which is only sensitive to the scale set by $z_{\text{cut}}$ since all of its emissions fail soft drop. $S_C(z_{\text{cut}} e_{2,L}^{(\alpha)})$ is a soft function that is boosted along the direction of the jet in the left hemisphere; its corresponding modes are referred to as collinear-soft
Figure 3.1: Schematic of the modes in the factorization theorem for soft-drop groomed hemispheres in $e^+e^- \rightarrow$ dijets events. $S_G(z_{\text{cut}})$ denotes the soft wide-angle modes, $S_C(z_{\text{cut}}e_2^{(\alpha)})$ denotes the collinear-soft modes, and $J(e_2^{(\alpha)})$ denotes the jet modes.
[57, 68, 127, 250, 251, 287]. Emissions in $S_C(z_{\text{cut}}^{(\alpha)}e_{2,L}^{(\alpha)})$ may or may not pass the soft drop requirement and are therefore constrained by both $z_{\text{cut}}$ and $e_{2,L}^{(\alpha)}$. Importantly, this collinear-soft mode depends on only a single scale which we generically denote by $z_{\text{cut}}e_{2,L}^{(\alpha)}$. (For $\alpha \neq 2$ or $\beta > 0$, the single scale is a different combination of $z_{\text{cut}}$ and $e_{2,L}^{(\alpha)}$; we simply call it $z_{\text{cut}}e_{2,L}^{(\alpha)}$ for notational brevity.) $J(e_{2,L}^{(\alpha)})$ is the jet function for the left hemisphere jet, and all emissions in the jet function parametrically pass the soft drop requirement. Thus, the jet function is independent of the scale set by $z_{\text{cut}}$, and only depends on $e_{2,L}^{(\alpha)}$. $\otimes$ denotes convolution in $e_{2,L}^{(\alpha)}$, and a similar collinear-soft and jet factorization exists for the right hemisphere.

As we will explain in detail, there are several important consequences of this factorization formula. Because the formula depends on the observables $e_{2,L}^{(\alpha)}, e_{2,R}^{(\alpha)}$ only through collinear objects each of which has a single scale, there are no non-global logarithms. The elimination of the purely soft contribution also makes the shape of soft-drop groomed jet shapes largely independent of what else is going on in the event. For example, the shape of the left hemisphere jet mass is independent of what is present in the right hemisphere. Additionally, the scale associated with the collinear-soft mode is parametrically larger than the soft scale associated with ungroomed masses, so non-perturbative corrections such as hadronization are correspondingly smaller.

This factorization theorem allows us to go beyond NLL accuracy to arbitrary accuracy. In this paper, we show that next-to-next-to-leading logarithmic (NNLL) accu-
racy is readily achievable. We focus on $\alpha = 2$ where the two-point energy correlation function is equal to the squared jet mass (up to a trivial normalization). This lets us extract most of the necessary two-loop anomalous dimensions from the existing literature. For $\beta = 0$, the global soft function $S_G(z_{\text{cut}})$ is closely related to the soft function with an energy veto [127, 309] which is known to two-loop order. There are additional clustering effects from the soft drop algorithm, but these are straightforward to calculate. Interestingly, we find that the clustering effects in the soft drop groomer are intimately related to similar effects observed in jet veto calculations [46, 49, 65, 67, 302]. For $\beta = 1$, we compute the two-loop anomalous dimension of $S_G(z_{\text{cut}})$ numerically using the fixed-order code EVENT2 [113]. We thereby achieve full NNLL resummation for the soft-drop groomed jet mass.\(^2\)

It is straightforward to generalize from $e^+e^-$ to $pp$ collisions, since the distribution is determined by collinear physics within the jet, independent of the initial state. The main new ingredient in $pp$ collisions is that jets may be initiated by quarks or gluons. As we will show, soft-drop grooming the jet enables an infrared and collinear safe definition of the jet flavor at leading power in $e_2^{(\alpha)}$ and $z_{\text{cut}}$ by simply summing the

\(^1\)While we will not do it in this paper, one could use the results of Ref. [71] which calculates the anomalous dimension of the soft function for event-wide (recoil-free) angularities [30, 76, 170, 252] or energy correlation functions with arbitrary angular exponent. This would enable us to extend our results to the case with $\alpha \neq 2$.

\(^2\)The jet mass has been calculated at NNLL using other methods [128, 149, 224] as has 2-subjectivity [176]. However, without grooming the jets, there are non-global logarithms which are not resummed (and which may or may not be quantitatively important) and uncontrollable sensitivity to pileup (which is very quantitatively important).
flavors of partons in the groomed jet. Using this procedure, we are able to match our
NNLL resummed distribution of soft-drop groomed jet mass to fixed order results for
\( pp \rightarrow Z + j \) events (including relative \( \mathcal{O}(\alpha_s^2) \) corrections to the Born process).

The outline of this paper is as follows. In Sec. 3.2, we review the definition of the
soft drop grooming algorithm and the energy correlation functions. In Sec. 3.3, we
present the factorization theorem for soft-drop groomed energy correlation functions
in \( e^+e^- \rightarrow \text{dijets} \) events. In this section, we will also present a detailed power-counting
analysis of soft-dropped observables to determine the range of validity of the factoriza-
tion theorem. Our factorization theorem has many non-trivial consequences, which
we review in Sec. 3.4. These include absence of non-global logarithms, process inde-
pendence, and small hadronization corrections. In Sec. 3.5, we describe and present
the ingredients necessary for NNLL resummation. Here, we also describe our method
for extracting anomalous dimensions from EVENT2. We then match our NNLL results
with fixed-order calculations for \( e^+e^- \) collisions in Sec. 3.6 and for \( pp \rightarrow Z + j \) events
in Sec. 3.7, comparing with Monte Carlo simulations in each case. In Sec. 3.8, we sum-
marize and conclude. The calculational details for NNLL resummation are collected
in appendices.
3.2 Observables

In this section, we review the soft drop grooming algorithm and the energy correlation functions. Although previous work has focused on jets produced in $pp$ collisions, we will provide definitions for both lepton and hadron collider environments.

3.2.1 Soft Drop Grooming Algorithm

Given a set of constituents of a jet with radius $R$, the soft drop grooming algorithm [248] proceeds in the following way:

1. Recluster the jet with a sequential $k_T$-type [110, 111, 168] jet algorithm. This produces an infrared and collinear (IRC) safe branching history of the jet. The $k_T$ clustering metric for jets in $e^+e^-$ collisions is

   \[ d_{ij}^{e^+e^-} = \min \left[ E_i^{2p}, E_j^{2p} \right] (1 - \cos \theta_{ij}), \]

   (3.2.1)

   where $E_i, E_j$ are the energies of particles $i$ and $j$ and $\theta_{ij}$ is their relative angle. $p$ is a real number that defines the particular jet algorithm. For jets produced in $pp$ collisions, the $k_T$ clustering metric is

   \[ d_{ij}^{pp} = \min \left[ p_{T_i}^{2p}, p_{T_j}^{2p} \right] R_{ij}^2, \]

   (3.2.2)

   where $p_{T_i}, p_{T_j}$ are the transverse momenta of particles $i$ and $j$ with respect to the beam and $R_{ij}^2$ is their relative angle in the pseudorapidity-azimuth angle plane.
While the original implementation of soft drop was restricted to reclustering with the Cambridge/Aachen algorithm \( (p = 0) \) [163, 311, 312], we will also briefly consider reclustering with the anti-\( k_T \) algorithm \( (p = -1) \) [97] in Sec. 3.5.2.

2. Sequentially step through the branching history of the reclustered jet. At each branching, check the soft drop criterion. For \( e^+ e^- \) collisions, we require

\[
\frac{\min[E_i, E_j]}{E_i + E_j} > z_{\text{cut}} \left( \sqrt{2} \frac{\sin \theta_{ij}}{\sin \frac{\theta_T}{2}} \right)^\beta.
\]

This is known as the soft drop criterion. If the branching fails this requirement, then the softer of the two daughter branches is removed from the jet.

The soft drop groomer then continues to the next branching in the remaining clustering history. For \( pp \) collisions, the soft drop criterion is

\[
\frac{\min[p_{T_i}, p_{T_j}]}{p_{T_i} + p_{T_j}} > z_{\text{cut}} \left( \frac{R_{ij}}{R} \right)^\beta.
\]

3. The procedure continues until the soft drop criterion is satisfied. At that point, soft drop terminates, and returns the jet groomed of the branches that failed the soft drop criterion.

Once the jet has been groomed, any observable can be measured on its remaining constituents.
3.2.2 Energy Correlation Functions

On jets that have been groomed by soft drop, we measure the two-point energy correlation functions [47, 221, 253]. We do this mainly for concreteness; the general properties of the factorized formula we will present apply for a much broader class of observables. For jets in $e^+e^-$ collisions, the two-point energy correlation function $e_2^{(\alpha)}$ is

$$
e_2^{(\alpha)}|_{e^+e^-} = \frac{1}{E_J^2} \sum_{i<j \in J} E_i E_j \left( \frac{2p_i \cdot p_j}{E_i E_j} \right)^{\alpha/2}, \tag{3.2.5}$$

where $E_J$ is the sum of the energies of particles in the jet, the sum runs over distinct pairs $i, j$ of particles in the jet, $p_i$ is the four-vector momentum of particle $i$, and the angular exponent $\alpha$ is required to be greater than 0 for IRC safety. For $\alpha = 2$ and a jet that has massless constituents, the two-point energy correlation function reduces to the normalized, squared jet mass:

$$
e_2^{(2)}|_{e^+e^-} = \frac{m_J^2}{E_J^2}. \tag{3.2.6}$$

The energy correlation functions have the nice property that they are insensitive to recoil effects [47, 252] and do not include explicit axes in their definition.

For jets produced in $pp$ collisions, the energy correlation functions are appropriately modified by replacing spherical coordinates with cylindrical coordinates:

$$
e_2^{(\alpha)}|_{pp} = \frac{1}{p_{T,J}^2} \sum_{i<j \in J} p_{T,i} p_{T,j} R_{ij}^\alpha, \tag{3.2.7}$$

where $p_{T,J}$ is the transverse momentum of the jet and $R_{ij}$ is the separation of parti-
cles $i$ and $j$ in the pseudorapidity-azimuthal angle plane. For jets at central rapidities and in the limit that all particles in the jet are collinear, Eq. (3.2.7) reduces to Eq. (3.2.5). This property in particular will enable us to recycle results calculated in $e^+e^-$ collisions to the case of $pp$ collisions.

### 3.3 Factorization Theorem

In this section, we derive the factorization formula for energy correlation functions measured on soft-drop groomed jets in the region of phase space where $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$ using SCET [51, 52, 55, 56]. We begin with a power counting analysis based on the scales $e_2^{(\alpha)}$ and $z_{\text{cut}}$ relevant to soft-drop groomed jets. This enables us to identify all modes and their momentum scalings that contribute at leading power. Using these scales and the associated modes, we derive the factorization formula. We then show that the jet function in the factorization formula can be re-factorized due to a collinear-soft mode which decouples from the collinear-but-not-soft modes as a result of soft drop.

#### 3.3.1 Power Counting and Modes

For jets on which the soft drop groomer is applied and the energy correlation functions are measured, there are three relevant dimensionless scales: the jet radius $R$, the soft drop parameter $z_{\text{cut}}$, and $e_2^{(\alpha)}$. Typically, jet radii are $R \sim 1$. We are in-
Figure 3.2: Location of modes appearing in the soft drop factorization theorem in the plane defined by energy fraction $z$ and splitting angle $\theta$ of emissions in the jet. The solid diagonal line separates the regions of phase space where emissions pass and fail soft drop. All emissions along the dashed line that pass soft drop contribute at leading power to the measured value of $\epsilon_2^{(\alpha)}$. 
terested in the singular region $e_2^{(\alpha)} \to 0$ for a fixed value of $z_{\text{cut}}$. Thus we can assume $e_2^{(\alpha)} \ll z_{\text{cut}}$. We will also assume $z_{\text{cut}} \ll 1$ to refactorize the jet function. The limits $R \ll 1$ or $z_{\text{cut}} \sim 1$ could be considered as well, but are beyond the scope of our analysis.

We will use scaling arguments to identify the regions of phase space that are present at leading power and then take the limit where each region becomes a separate sector, that no longer interacts with the other regions.

For a jet to have $e_2^{(\alpha)} \ll 1$, all particles must be either soft or collinear to the jet axis. In particular, a particle with energy $E = zE_j$ at an angle $\theta$ from the jet axis must satisfy

$$z\theta^\alpha \lesssim e_2^{(\alpha)}.$$  \hspace{1cm} (3.3.1)

This is a line in the $\log(1/z)$-$\log(1/\theta)$ plane, as shown in Fig. 3.2. Anything below the dashed line in this figure is too hard to be consistent with a given value of $e_2^{(\alpha)}$. The soft drop criterion is that

$$z_{\text{cut}} \lesssim z\theta^{-\beta},$$  \hspace{1cm} (3.3.2)

This is the region below the solid line in Fig. 3.2.

To find the relevant modes for the factorized expression, we need to identify the distinct characteristic momentum scalings that approach the singular regions of phase space in the limit $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$. For a particular scaling, the constraints in Eqs. (3.3.1) and (3.3.2) will either remain relevant or decouple. We can characterize
the relevant regions by their scalings in light-cone coordinates. Defining $n^\mu$ as the jet
direction and $\tilde{n}^\mu$ as the direction backwards to the jet, then light-cone coordinates are
triplets $p = (p^-, p^+, p_\perp)$ where $p^- = \tilde{n} \cdot p$, $p^+ = n \cdot p$ and $p_\perp$ are the components
transverse to $n$. On-shell massless particles have $p^+p^- = p_\perp^2$. The energy fraction
is $z = p^0/Q = \frac{1}{2}(p^+ + p^-)/Q$ and the angle to the jet axis in the collinear limit is
$\theta = p_\perp/p^0$.

We start with the **soft modes**, emitted at large angles $\theta \sim 1$, but still within the
jet. If such radiation were to pass soft drop, with energy fraction greater than $z_{\text{cut}}$,
it would set $e_2^{(\alpha)} \gtrsim z_{\text{cut}}$; this contradicts our assumed hierarchy $e_2^{(\alpha)} \ll z_{\text{cut}}$. There-
fore, soft wide-angle radiation is removed by soft drop and is not constrained by $e_2^{(\alpha)}$.
These modes thus have momenta that scale like

$$p_s \sim z_{\text{cut}} Q(1,1,1).$$  \hspace{1cm} (3.3.3)

They contribute only to the normalization of the distribution, not to its shape.

Next consider the collinear radiation, emitted at small angles $\theta \ll 1$. All collinear
radiation has $p^- \gg p^+$. Then, from Eq. (3.3.1), we find

$$e_2^{(\alpha)} \sim \frac{(p^+)^{\alpha/2}(p^-)^{1-\alpha/2}}{Q}$$  \hspace{1cm} (3.3.4)

Collinear modes can either have $z \sim 1$ or be parameterically soft $z \ll 1$.

For modes with $z \sim 1$, we have $z \gg z_{\text{cut}}$. Thus $p^- \sim Q$ and $p^+ \sim Q(e_2^{(\alpha)})^{2/\alpha}$
independent of $z_{\text{cut}}$. Their scaling is

$$p_c \sim Q \left(1,(e_2^{(\alpha)})^{2/\alpha},(e_2^{(\alpha)})^{1/\alpha}\right).$$  \hspace{1cm} (3.3.5)

108
We call these modes **collinear modes**, although strictly they are not-soft collinear modes.

Collinear radiation that can have $z \sim z_{\text{cut}} \ll 1$ we call **collinear-soft**. In this case, $p^- \sim zQ$ and $p^+ \sim \theta^2 zQ$. These modes are simultaneously compatible with Eqs. (3.3.1) and (3.3.2). Their scaling is determined by saturating these parametric relationships, which leads to

$$ p_{cs} \sim (z_{\text{cut}})^{\frac{\alpha}{\alpha + \beta}} (e(2)^{(\alpha)}_{2})^{\frac{\beta}{\alpha + \beta}} Q \left( 1, \left( \frac{e(2)^{(\alpha)}_{2}}{z_{\text{cut}}} \right)^{\frac{2}{\alpha + \beta}}, \left( \frac{e(2)^{(\alpha)}_{2}}{z_{\text{cut}}} \right)^{\frac{1}{\alpha + \beta}} \right). $$ \hspace{1cm} (3.3.6)

This is the point in phase space labeled $S_C(z_{\text{cut}}e(2)^{(\alpha)}_{2})$ in Fig. 3.2.

### 3.3.2 Factorization and refactorization

With the relevant scalings identified, we proceed to derive the factorization formula.

For simplicity, we focus on the case of $e^+e^- \rightarrow$ hemisphere jets, with $e(2)^{(\alpha)}_{2}$ measured on each hemisphere. Jets at hadron colliders can be treated similarly, as we discuss in Sec. 3.4. Fig. 3.3 illustrates the relevant modes and their scales.

We begin with the usual SCET factorization formula, in the absence of soft drop grooming. The hard, collinear and soft modes are separated in the limit of small observables. This leads to [170, 179, 293]

$$ \frac{d^2\sigma}{dc_2^{(\alpha)} dc_2^{(\alpha)}} = H(Q^2) \times S \left( e_2^{(\alpha)}_{2,L}, e_2^{(\alpha)}_{2,R} \right) \otimes J(e_2^{(\alpha)}_{2,L}) \otimes J(e_2^{(\alpha)}_{2,R}) $$ \hspace{1cm} (3.3.7)

for the ungroomed hemispheres in $e^+e^- \rightarrow$ dijets events, provided $e_2^{(\alpha)}_{2,L}, e_2^{(\alpha)}_{2,R} \ll 1$.

Here, $\otimes$ denotes convolution in $e_2^{(\alpha)}_{2,L}$ or $e_2^{(\alpha)}_{2,R}$ appropriately. To get to this equation, one
Figure 3.3: Illustration of the multi-stage matching procedure to derive the soft drop factorization theorem. As discussed in the text, we first match QCD to SCET, then factorize the jet function into collinear and collinear-soft modes. Canonical scales of all modes in the factorization theorem are shown on the right, ordered in virtuality where we assume that $\alpha > 1$ and $\beta \geq 0$.

can match to full QCD to get the hard function, then decouple the soft and collinear degrees of freedom to pull the jet and soft functions apart [51, 52, 55, 56]. Alternatively, one can use the method of regions approach [60, 74], or the on-shell phase space approach [172, 174, 177]. Importantly, $e_2^{(\alpha)}$ is insensitive to recoil effects from soft emissions that displace the jet axis from the direction of hard, collinear particles [47, 252], and so the jet and soft functions are completely decoupled.

Next we write down the hard-soft-jet factorization formula in the presence of soft drop grooming, assuming the hierarchy $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$. With this assumption, soft radiation emitted at large angles must necessarily fail the soft drop criterion. Thus, all wide angle soft radiation in the jets (in this case, the hemisphere jets) is groomed and cannot contribute to the observable. All that remains of the global soft function
is a $z_{\text{cut}}$-dependent normalization factor $S_G(z_{\text{cut}})$. This leads to

$$
\frac{d^2 \sigma}{d e^{(\alpha)}_{2,L} d e^{(\alpha)}_{2,R}} = H(Q^2) \times S_G(z_{\text{cut}}) \times J_{ze}\left(z_{\text{cut}}, e^{(\alpha)}_{2,L}\right) \times J_{ze}\left(z_{\text{cut}}, e^{(\alpha)}_{2,R}\right). \quad (3.3.8)
$$

$S_G(z_{\text{cut}})$ gives the cross section for the radiation from a set of Wilson lines that fails the soft drop criterion. An explicit calculation of $S_G$ for hemisphere jets at one-loop is given in Appendix B.3. With the collinear and soft modes decoupled, we can lower the virtuality of the collinear modes without further matching.

The jet function $J_{ze}$ still depends on multiple scales, so to resum all the large logarithms it must be re-factorized. To see that it re-factorizes, note first that in addition to being collinear, radiation in the jet function that is sensitive to the scale set by $z_{\text{cut}}$ must also be soft, by the assumption that $z_{\text{cut}} \ll 1$. Equivalently, emissions with order-1 energy fractions are not constrained by the scale $z_{\text{cut}}$. We can thus factorize the jet function into two pieces depending on their energy fraction:

$$
J_{ze}(z_{\text{cut}}, e^{(\alpha)}_2) = J(e^{(\alpha)}_2) \otimes S_C(z_{\text{cut}}e^{(\alpha)}_2). \quad (3.3.9)
$$

Here, $J(e^{(\alpha)}_2)$ is the jet function that only depends on $e^{(\alpha)}_2$ and only receives contributions from emissions with order-1 energy fraction. $S_C(z_{\text{cut}}e^{(\alpha)}_2)$ is the soft limit of the unfactorized jet function $J_{ze}(z_{\text{cut}}, e^{(\alpha)}_2)$. The scaling of the collinear and collinear-soft modes are given in Eqs. (3.3.5) and (3.3.6) as discussed above. Note that, importantly, because the collinear-soft mode arises from re-factorization of a jet function, it is a color singlet and only depends on two back-to-back directions. Because the jet function only depends on $e^{(\alpha)}_2$, it is sensitive to a single infrared scale.
The step in Eq. (3.3.9) is the most unusual and important in the derivation. That the collinear-soft function depends on only a single combination of $z_{\text{cut}}$ and $e^{(\alpha)}_2$ is absolutely critical to being able to resum all the logs of $e^{(\alpha)}_2$. We therefore devote Sec. 3.3.3 to showing explicitly that the collinear-soft function depends on a unique combination of $z_{\text{cut}}$ and $e^{(\alpha)}_2$ as determined by the parametric scaling of the modes of Eq. (3.3.6), and so is also only sensitive to a single infrared scale.

Inserting Eq. (3.3.9) into Eq. (3.3.8) results in the factorization formula for soft drop energy correlation functions:

$$\frac{d^2 \sigma}{d e^{(\alpha)}_2 d e^{(\alpha)}_{2,R}} = H(Q^2) S_G(z_{\text{cut}}) \left[ S_C \left( z_{\text{cut}} e^{(\alpha)}_{2,L} \right) \otimes J(e^{(\alpha)}_{2,L}) \right] \left[ S_C \left( z_{\text{cut}} e^{(\alpha)}_{2,R} \right) \otimes J(e^{(\alpha)}_{2,R}) \right]$$

(3.3.10)

The pieces of the factorization theorem are:

- $H(Q^2)$ is the hard function for production of an $e^+e^- \to \text{dijets}$ event.

- $S_G(z_{\text{cut}})$ is the global soft function. It integrates the radiation coming from Wilson lines in the jet directions that fails the soft drop criterion. Its modes fail soft drop and have momenta that scale as determined in Eq. (3.3.3).

- $J(e^{(\alpha)}_2)$ is the jet function describing the emission of collinear radiation from a jet. Its modes parametrically pass soft drop and have momenta that scale as determined in Eq. (3.3.5).

- $S_C \left( z_{\text{cut}} e^{(\alpha)}_2 \right)$ is the collinear-soft function describing the emission of soft radi-
ation boosted along the direction of a jet. Its modes may or may not pass soft
drop and have momenta that scale as determined in Eq. (3.3.6). We denote the
single scale that the collinear-soft function depends on as $z_{\text{cut}} e_{2,L}^{(a)}$ for brevity; it
is shorthand for Eq. (3.3.27).

We present the operator definitions and explicit one-loop results for all of these func-
tions in the appendices.

The appearance of collinear-soft modes in this factorization theorem has some
similarities and differences with respect to the identification of other collinear-soft
modes in the literature [57, 68, 127, 250, 251, 287]. The original construction of a
collinear-soft mode in Ref. [57] followed from boosting two jets in an event far from
their center-of-mass frame in an effective theory of collinear dijets called SCET$_+$. The
collinear-soft mode in SCET$_+$ is sensitive to three Wilson line directions: the two
from the collinear jets and the backward direction from boosting all other jets in the
event. This collinear-soft mode was also exploited in Ref. [250] in the resummation of
jet observables that are sensitive to multi-prong substructure.

The collinear-soft mode in the factorization theorem presented here, however, is
more similar to modes identified from the measurement of multiple observables on a
jet, each of which is only sensitive to radiation about a single hard core [68, 127, 250,
251, 287]. For example, Ref. [287] presented a factorization theorem for jets on which
two angularities [30, 76, 170, 252] are measured. At leading power, angularities are
only sensitive to the hard jet core, and so the collinear-soft modes only know about two Wilson line directions: the jet axis and the backward direction. More recently, collinear-soft modes of this type have been used to resum NGLs \cite{68, 251} and logarithms of the jet radius \cite{127}.

### 3.3.3 The Single Scale of the Collinear-Soft Function

To demonstrate explicitly that the collinear-soft function only depends on a single scale, we can make the following scaling argument. The collinear-soft function has the following form:

\[
S_C \left( z_{\text{cut}}, e_{2}^{(\alpha)} \right) = \sum_{n} \mu^{2n} \int d\Pi_{n} |\mathcal{M}_{n}|^{2} \Theta_{\text{SD}} \delta_{e_{2}^{(\alpha)}}. \tag{3.3.11}
\]

Here, \( n \) is the number of final state collinear-soft particles, \( d\Pi_{n} \) is on-shell Lorentz-invariant phase space in \( d = 4 - 2\epsilon \) dimensions:

\[
d\Pi_{n} = \prod_{i=1}^{n} \frac{d^{d}k_{i}}{(2\pi)^{d}} 2\pi \delta(k_{i}^{2}) \Theta(k_{i}^{0}), \tag{3.3.12}
\]

\( \mu \) is the renormalization scale, and \( \mathcal{M}_{n} \) is the amplitude for the production of the final state. \( \Theta_{\text{SD}} \) represents the soft drop grooming algorithm, which applies constraints on the final state and \( \delta_{e_{2}^{(\alpha)}} \) represents the measurement of \( e_{2}^{(\alpha)} \) on the final state:

\[
\delta_{e_{2}^{(\alpha)}} = \delta \left( e_{2}^{(\alpha)} - \frac{2\alpha}{Q} \sum_{i} (k_{i}^{-})^{1-\alpha/2}(k_{i}^{+})^{\alpha/2} \right), \tag{3.3.13}
\]

where the sum runs over the set of final state particles \( \{i\} \) that remain in the jet after grooming. To write this expression, we have used the definition of \( e_{2}^{(\alpha)} \) from Sec. 3.2.2.
and expanded in the collinear-soft limit, as in Eq. (3.3.4).

Now, we rescale the momenta in light-cone coordinates that appear in the phase space integral in the following way:

\[
k^− \to (z_{\text{cut}})^{- \frac{a}{a+\beta}} \left( \epsilon_2^{(\alpha)} \right)^{\frac{\alpha}{a+\beta}} k^−, \tag{3.3.14}
\]

\[
k^+ \to (z_{\text{cut}})^{- \frac{a-2}{a+\beta}} \left( \epsilon_2^{(\alpha)} \right)^{\frac{2+\beta}{a+\beta}} k^+, \]

\[
k_\perp \to (z_{\text{cut}})^{- \frac{a-1}{a+\beta}} \left( \epsilon_2^{(\alpha)} \right)^{\frac{1+\beta}{a+\beta}} k_\perp.
\]

At leading power in exactly \( d = 4 \), the phase space measure \( d\Pi_n \) and the squared matrix element \( |\mathcal{M}_n|^2 \) scale exactly inversely. Therefore, in \( d \) dimensions, under this rescaling, we have

\[
d\Pi_n |\mathcal{M}_n|^2 \to \left( (z_{\text{cut}})^{- \frac{a-1}{a+\beta}} \left( \epsilon_2^{(\alpha)} \right)^{\frac{1+\beta}{a+\beta}} \right)^{-2^n e} d\Pi_n |\mathcal{M}_n|^2. \tag{3.3.15}
\]

Next, look at how the measurement functions \( \Theta_{\text{SD}} \) and \( \delta_{\epsilon_2^{(\alpha)}} \) change under the rescaling of Eq. (3.3.14). First, consider the soft drop groomer \( \Theta_{\text{SD}} \). This consists of two parts: one, the reclustering with the Cambridge/Aachen algorithm and the second, the energy requirement on the clustered particles. The clustering metric of the Cambridge/Aachen algorithm is just the pairwise angle

\[
d_{ij}^{C/A} = \theta_{ij}^2, \tag{3.3.16}
\]

and a pair \( \{i, j\} \) of particles in the jet are clustered if they have the smallest \( d_{ij}^{C/A} \).

Importantly, the reclustering of the jet with soft drop is completely inclusive: all particles in the jet are clustered with no jet radius parameter. Therefore, for collinear-
soft modes, there are only three types of clustering constraints that can be enforced, depending on what $d_{ij}^{C/A}$’s are being compared. If in the clustering history we compare the angles between two collinear-soft particles $i$ and $j$ to the jet axis, this corresponds to the constraint
\[
\Theta \left( \frac{k_i^+}{k_i^-} - \frac{k_j^+}{k_j^-} \right),
\] (3.3.17)
This is invariant under the rescalings of Eq. (3.3.14). If in the clustering history we compare the angle between a collinear-soft particle $i$ to the jet axis and the angle between two collinear-soft particles $j$ and $k$, we have the constraint
\[
\Theta \left( \frac{k_i^+}{k_i^-} - \frac{k_j \cdot k_k}{k_j^- k_k^-} \right),
\] (3.3.18)
which is also invariant under the rescalings of Eq. (3.3.14). Finally, we can compare the angle between a pair of collinear-soft particles $i$ and $j$ to the angle between another pair of collinear-soft functions $k$ and $l$, this leads to
\[
\Theta \left( \frac{k_i \cdot k_j}{k_i^- k_j^-} - \frac{k_k \cdot k_l}{k_k^- k_l^-} \right).
\] (3.3.19)
This too is invariant under Eq. (3.3.14). Therefore, for all possible clustering structures, the Cambridge/Aachen algorithm is invariant under the rescalings of Eq. (3.3.14).

The soft drop energy requirement on any number of particles that have been reclustered takes the form:
\[
\Theta \left( \sum_i z_i - z_{\text{cut}} \theta^\beta \right),
\] (3.3.20)
where $z_i$ is the energy fraction of particle $i$ and $\theta$ is the angle that the cluster of par-
articles \{i\} makes with the jet axis. In terms of light-cone coordinates, this can be written as:

\[
\Theta \left( \sum_i z_i - z_{\text{cut}} \theta^\beta \right) = \Theta \left( k^- - z_{\text{cut}} Q \left( \frac{k_\perp}{k^-} \right)^\beta \right),
\]

where

\[
k_\perp = \left| \sum_i \tilde{k}_{\perp,i} \right|,
\]

\[
k^- = \sum_i k_i^-.
\]

Applying the rescalings of Eq. (3.3.14), this constraint becomes

\[
\Theta \left( \sum_i z_i - z_{\text{cut}} \theta^\beta \right) \rightarrow \Theta \left( \sum_i z_i - \theta^\beta \right).
\]

Note that the low scale \(z_{\text{cut}}\) has been removed from this constraint.

Under the rescaling, the measurement constraint \(\delta_{e_2}^{(\alpha)}\) becomes

\[
\delta \left( e_2^{(\alpha)} - \frac{2 \alpha}{Q} \sum_i (k_i^-)^{1-\alpha/2}(k_i^+)^{\alpha/2} \right) \rightarrow \frac{1}{e_2^{(\alpha)}} \delta \left( 1 - \frac{2 \alpha}{Q} \sum_i (k_i^-)^{1-\alpha/2}(k_i^+)^{\alpha/2} \right).
\]

Therefore, the low scale \(e_2^{(\alpha)}\) has been removed from this constraint.

Putting this all together, the collinear-soft function can be rewritten as

\[
S_C \left( z_{\text{cut}}, e_2^{(\alpha)} \right) = \sum_n \mu^{2n} \left( \frac{z_{\text{cut}}}{\pi^{n+1}} \left( e_2^{(\alpha)} \right)^{1+\beta} \right)^{-2n} \int d\Pi_n |\mathcal{M}_n|^2 \Theta_{\text{SD}}^{z_{\text{cut}}=1} \delta_{e_2^{(\alpha)}}^{1}.
\]

We have used the notation that \(\Theta_{\text{SD}}^{z_{\text{cut}}=1}\) is the soft drop grooming algorithm with \(z_{\text{cut}} = 1\) and \(\delta_{e_2^{(\alpha)}}^{1}\) is the measurement with \(e_2^{(\alpha)} = 1\). All low scales have been explicitly removed from the phase space integral. This function is now seen to be a function
only of the single scale

\[ z_{\text{cut}} e_2^{(\alpha)} = (z_{\text{cut}})^{\frac{\alpha - 1}{\alpha + 3}} (e_2^{(\alpha)})^{\frac{1 + \beta}{\alpha + 3}}. \]  

(3.3.27)

The quotes just mean that the left-hand side is our abbreviation for the unwieldly quantity on the right-hand side.

This proves that, to all orders, the collinear-soft function has dependence on only a single infrared scale, defined by this combination of \( z_{\text{cut}} \) and \( e_2^{(\alpha)} \). Notice that the proof relied on the choice of Cambridge/Aachen reclustering in the soft drop grooming algorithm.

This completes the derivation of factorization and re-factorization for soft-drop groomed hemispheres in \( e^+e^- \) collisions, on which two-point energy correlation functions have been measured. All functions in the factorized cross section in Eq. (3.3.10) are sensitive to a single infrared scale and so all large logarithms can be resummed with the renormalization group.

### 3.4 Consequences of Factorization Theorem

Before we use the factorization theorem of Eq. (3.3.10) to make predictions for the cross section, we discuss consequences of this formula in some detail. Because the factorization theorem was derived without respect to any fixed order, these results hold to all orders.

Many of these consequences follow from the fact that soft wide angle radiation does
not contribute to the shape of the soft-drop groomed $e_2^{(a)}$ distribution for $e_2^{(a)} \ll z_{\text{cut}} \ll 1$, a property that persists even for jets at hadron colliders. For example, it follows immediately from this fact that the shape of such a distribution is insensitive to contamination from pile-up and underlying event.

In this section, we will furthermore prove that at leading power, there are no NGLs that affect the shape of the soft-drop groomed $e_2^{(a)}$ distribution in this regime. This was explicitly shown at $O(\alpha_s^2)$ in Refs. [147, 148, 248] and plausibility arguments were presented for all orders, but this is the first proof. The factorization theorem also exhibits sample independence to a large degree, because the shape of the distribution is only sensitive to collinear physics. We will also demonstrate that soft-drop groomed energy correlation functions are less sensitive to hadronization than their ungroomed counterparts.

### 3.4.1 Absence of Non-Global Logarithms to All Orders

NGLs in cross sections of observables measured on individual hemispheres in $e^+e^-$ collisions arise from a parametric separation of the scales in the hemispheres. Their leading effects are exclusively non-Abelian and quantify the correlation between the two hemispheres. Clearly, for a correlation to be present, there must be correlated radiation emitted into both hemispheres. If we measure the energy correlation functions $e_2^{(a)}$ on both hemispheres and demand that $e_2^{(a)} \ll 1$, then the radiation in the event
must be soft wide-angle or collinear. At leading power, it is not possible to have correlations between different collinear directions (beyond total momentum conservation) as this would violate the collinear factorization of gauge theory amplitudes. Therefore, correlations and NGLs can only arise from soft, wide-angle radiation in the event with these assumptions.

The factorization theorem for ungroomed hemisphere energy correlation functions is \[170, 179, 293\]

\[
\frac{d^2 \sigma}{de_{2,L} de_{2,R}} \bigg|_{\text{ug}} = H(Q^2) S(e_{2,L}^{(\alpha)}, e_{2,R}^{(\alpha)}) \otimes J(e_{2,L}^{(\alpha)}) \otimes J(e_{2,R}^{(\alpha)}),
\]

where “ug” denotes ungroomed. The cross section explicitly depends on soft wide-angle radiation through the soft function \(S(e_{2,L}^{(\alpha)}, e_{2,R}^{(\alpha)})\), and so if either \(e_{2,L}^{(\alpha)} \ll e_{2,R}^{(\alpha)}\) or \(e_{2,R}^{(\alpha)} \ll e_{2,L}^{(\alpha)}\), NGLs will be present in this factorization theorem. Because the soft function depends on two scales, all of the singular dependence cannot be determined by renormalization group invariance. More generally, non-global structure present in the soft function has been studied at \(O(\alpha_s^2)\) \[218, 219, 227\] and beyond \[234, 294\] and recently, methods have been developed to control all-orders behavior \[68, 107, 251, 276\]. However, NGLs represent an obstruction to resummation of the cross section to NLL and beyond.

For groomed hemisphere energy correlation functions, our factorization theorem
instead takes the form of Eq. (3.3.10):

\[
\frac{d^2 \sigma}{de^{(α)}_{2,L} de^{(α)}_{2,R}} = H(Q^2) S_G(z_{\text{cut}}) \left[ S_C(z_{\text{cut}} e^{(α)}_{2,L}) \otimes J(e^{(α)}_{2,L}) \right] \left[ S_C(z_{\text{cut}} e^{(α)}_{2,R}) \otimes J(e^{(α)}_{2,R}) \right].
\]  

(3.4.2)

All soft, wide angle radiation throughout the event is described by \( S_G(z_{\text{cut}}) \), which is sensitive only to the single scale \( z_{\text{cut}} \). Therefore, there are no NGLs present in this factorization theorem. Even with a hierarchy between \( e^{(α)}_{2,L} \) and \( e^{(α)}_{2,R} \), these observables are completely decoupled at leading power in \( e^{(α)}_{2} \) and \( z_{\text{cut}} \). Additionally, the shape of the distribution is also independent of jet radius effects and the precise way in which the hemispheres are defined.

When we discuss soft-drop groomed jets in \( pp \) collisions in Sec. 3.7, we will place no constraint on global soft radiation throughout the event, unlike the case of \( e^+e^- \rightarrow \) hemisphere jets. Nevertheless, the shape of the soft-drop groomed \( e^{(α)}_{2} \) distribution will still have no NGLs, jet radius effects, etc., due to universality of the collinear limit of QCD amplitudes. The normalization, however, will in general be sensitive to scales both in the jet (set by \( z_{\text{cut}} \) and the jet radius \( R \)) and scales outside of the jet (set by the partonic collision energy). To eliminate these effects in \( pp \) collisions, we can normalize the cross section, say, to integrate to unity.
3.4.2 Process Independence

Strictly speaking, the factorization theorem of Eq. (3.3.10) depends on the process. It includes the hard function, which is process dependent, and a soft function, that knows about all hard jet directions. Nevertheless, there is a sense in which the factorization theorem is process independent. Normalizing the cross section completely removes the hard and soft function dependence. Then, by the universal collinear factorization of QCD amplitudes, if we are completely inclusive over the right hemisphere, then the differential cross section of the soft-drop groomed energy correlation function in the left hemisphere is given by

\[
\frac{d\sigma}{d\epsilon_{2,L}^{(\alpha)}} = N S_C(z_{\text{cut}} \epsilon_{2,L}^{(\alpha)}) \otimes J(\epsilon_{2,L}^{(\alpha)}),
\]

where we assume that \( \epsilon_{2,L}^{(\alpha)} \ll z_{\text{cut}} \ll 1 \) and \( N \) is some normalization factor. That is, in the deep infrared where \( \epsilon_{2,L}^{(\alpha)} \ll z_{\text{cut}} \ll 1 \), all radiation in the groomed jet is constrained to be collinear. Therefore, in this limit and for a fixed jet energy, the shape of the distribution for quark jets is independent of the process that created the quark jets, due to the universality of QCD matrix elements in the collinear limit.

This collinear factorization property of soft-drop groomed observables can be exploited for jets in \( pp \) collisions. Unlike the dominant case in \( e^+e^- \) collisions, jets at a \( pp \) collider can be either quark or gluon. Of course, on a jet-by-jet level, we cannot determine whether a jet was initiated by a quark or gluon. However, for a given process, we can determine the relative fraction of quark and gluon jets in the sample. For
jets produced at a \( pp \) collider, the process independence manifests itself in the cross section as

\[
\frac{d\sigma^{pp}}{de_2^{(\alpha)}} = D_q S_{C,q}(z_{\text{cut}} e_2^{(\alpha)}) \otimes J_q(e_2^{(\alpha)}) + D_g S_{C,g}(z_{\text{cut}} e_2^{(\alpha)}) \otimes J_g(e_2^{(\alpha)}),
\]

where \( D_q \) (\( D_g \)) is proportional to the fraction of quark (gluon) jets in the sample. The relative fraction of quark and gluon jets can be determined from fixed-order calculations, using a simple algorithm for determining the flavor of a groomed jet. We will describe this in detail in Sec. 3.7 when we match our resummed distribution to fixed order in \( pp \to Z + j \) events.

### 3.4.3 Hadronization Corrections

With a factorization formula, one can estimate the size and importance of non-perturbative corrections to the cross section. We will only consider non-perturbative corrections to the shape, as the normalization can be set by hand. Therefore, non-perturbative corrections can only enter into our factorization theorem via the jet or collinear-soft functions.

From Eqs. (3.3.5) and (3.3.6), the scales appearing in the jet and collinear-soft functions are

\[
\mu_J = Q \left( e_2^{(\alpha)} \right)^{1/\alpha},
\]

\[
\mu_{SC} = Q \left( z_{\text{cut}}^{\frac{\alpha + \beta}{\alpha + \beta}} e_2^{(\alpha)} \right)^{\frac{1 + \beta}{\alpha + \beta}}.
\]

If either of the scales approaches \( \Lambda_{\text{QCD}} \), then we expect there to be large corrections
to the perturbative cross section due to non-perturbative physics. We can estimate
when non-perturbative corrections become large by setting these scales to be $\Lambda_{\text{QCD}}$.

For $\alpha > 1$ and $\beta \geq 0$, the collinear-soft mode has a lower virtuality than the collinear
mode, so it will probe the non-perturbative region of phase space first. The value of
$e_2^{(\alpha)}$ at which the collinear-soft mode becomes non-perturbative is

$$
\mu_{SC} = \Lambda_{\text{QCD}} \implies e_2^{(\alpha)}|_{\text{NP}} \simeq \left( \frac{\Lambda_{\text{QCD}}}{z_{\text{cut}}Q} \right)^{\alpha-1} \frac{\Lambda_{\text{QCD}}}{Q}.
$$

(3.4.7)

This estimate can be compared with the Monte Carlo analysis of hadronization cor-
rections to the soft-drop groomed energy correlation functions from Ref. [248]. In
particular, the estimate of Eq. (3.4.7) of when non-perturbative corrections become
important for $\alpha = 2$ as a function of $\beta$ agrees exceptionally well with Fig. 10(a) of
Ref. [248].

For $\beta < \infty$, the soft drop groomer reduces the effect of non-perturbative corrections
with respect to the ungroomed observable. This can be simply seen from Eq. (3.4.7),
in which the prefactor

$$
\left( \frac{\Lambda_{\text{QCD}}}{z_{\text{cut}}Q} \right)^{\alpha-1} \frac{1}{\alpha+\beta}
$$

(3.4.8)

approaches unity as the grooming is removed ($\beta \to \infty$). This factor is less than 1 for
$\beta < \infty$, provided $\alpha > 1$ and $\Lambda_{\text{QCD}} < z_{\text{cut}}Q$. For high energy jets, this suppression
can be substantial. For example, for $\alpha = 2$ (corresponding to jet mass) and $\beta = 0$
(corresponding to mMDT groomer) non-perturbative effects become important at
\[ e^{(2)}_2 \bigg|_{\beta=0} \simeq \frac{\Lambda_{QCD}^2}{z_{\text{cut}} Q^2}. \] (3.4.9)

This agrees with the estimate of the size of nonperturbative corrections for the mMDT groomer from Ref. [148].

### 3.5 Achieving NNLL Accuracy

In this section, we determine the anomalous dimensions necessary to resum the large logarithms of soft-dropped energy correlation functions through NNLL accuracy. The practical details of how one assembles these ingredients, in the framework of SCET, to construct a resummed cross section are given in App. B.6. We will discuss matching to fixed-order and demonstrate our ability to make phenomenological predictions in subsequent sections.

Resummation in SCET is accomplished with renormalization group evolution. Solving the renormalization group equations to a given logarithmic accuracy requires anomalous dimensions to a particular fixed order. The anomalous dimensions of the functions in the factorization theorem must sum to zero, because the cross section is independent of the renormalization scale.

Recall that, for $e^+e^- \rightarrow$ hemisphere jets, the factorization theorem for soft-drop
Table 3.1: $\alpha_s$-order of ingredients needed for resummation to the accuracy given. $\Gamma_{\text{cusp}}$ is the cusp anomalous dimension, $\gamma$ is the non-cusp anomalous dimension, and $\beta$ is the QCD $\beta$-function. $\tilde{F}(\partial_\omega)$ are the logarithms in the low-scale matrix elements that have been Laplace transformed and $c_F$ are constants in the low-scale matrix elements. The final column shows the relative order to which the resummed cross section can be matched to fixed-order.

groomed energy correlation functions is

$$
\frac{d^2\sigma}{de_{2,L}^{(a)}de_{2,R}^{(a)}} = H(Q^2)S_G(z_{\text{cut}}) \left[ S_C(z_{\text{cut}}e_{2,L}^{(a)}) \otimes J(e_{2,L}^{(a)}) \right] \left[ S_C(z_{\text{cut}}e_{2,R}^{(a)}) \otimes J(e_{2,R}^{(a)}) \right]. \quad (3.5.1)
$$

Table 3.1 presents the order to which anomalous dimensions and constants of the functions in this factorization theorem must be computed for particular logarithmic accuracy (see, e.g., Ref. [29]). The cusp anomalous dimension $\Gamma_{\text{cusp}}$ and the QCD $\beta$-function are known through three-loop order [75, 239, 246, 273, 304, 308] and we present them in App. B.1. The hard function $H(Q^2)$ for $e^+e^- \rightarrow q\bar{q}$ is known to high orders and its non-cusp anomalous dimension $\gamma_H$ is known at three-loop order [266, 307]; we present the relevant pieces in App. B.2. For arbitrary angular exponents $\alpha$ and $\beta$, little else in the factorization theorem is known at sufficiently high accuracy to resum to NNLL.

The goal of this section is to fill in the rest of the table, to achieve full NNLL accuracy. We start in Sec. 3.5.1 restricting to $\alpha = 2$ (jet mass) and $\beta = 0$ (mMDT

<table>
<thead>
<tr>
<th>$\Gamma_{\text{cusp}}$</th>
<th>$\gamma$</th>
<th>$\beta$</th>
<th>$\tilde{F}(\partial_\omega)$</th>
<th>$c_F$</th>
<th>Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>$\alpha_s$</td>
<td>$\alpha_s$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NLL</td>
<td>$\alpha_s^2$</td>
<td>$\alpha_s$</td>
<td>$\alpha_s^2$</td>
<td>$\alpha_s$</td>
<td>$\alpha_s$</td>
</tr>
<tr>
<td>NNLL</td>
<td>$\alpha_s^3$</td>
<td>$\alpha_s^2$</td>
<td>$\alpha_s^3$</td>
<td>$\alpha_s^2$</td>
<td>$\alpha_s$</td>
</tr>
</tbody>
</table>
groomer). For this case, all of the missing ingredients can be determined by recycling results from the literature, up to calculable clustering effects from the soft drop algorithm. In Sec. 3.5.3, we consider $\alpha = 2$ and $\beta \geq 0$ and demonstrate that one can extract unknown two-loop non-cusp anomalous dimensions with EVENT2. It is possible to extend our analysis to angular exponents for the energy correlation functions beyond $\alpha = 2$, but we do not do it in this paper.$^3$

3.5.1 NNLL for $\alpha = 2$, $\beta = 0$

We first consider angular exponents $\alpha = 2$ and $\beta = 0$. In this case, the soft drop requirement enforced at every branching reduces to an energy cut

$$\min[E_i, E_j] > z_{\text{cut}}(E_i + E_j). \quad (3.5.2)$$

On the soft-drop groomed jets we then measure

$$e_2^{(2)} = \frac{m_g^2}{E_g^2}, \quad (3.5.3)$$

where the subscript $g$ denotes that the mass and energy are measured on the groomed jet. The jet functions in the factorization theorem are independent of the soft drop groomer, so we are able to use results from the literature for these. The inclusive jet function has been calculated to two loops $[54, 59, 63, 91]$ and the non-cusp anomalous

$^3$The two-loop non-cusp anomalous dimension of the soft function for event-wide (recoil-free) angularities $[30, 76, 170, 252]$ as a function of the angular exponent has been calculated in Ref. [71]. Recoil-free angularities and two-point energy correlation functions have identical anomalous dimensions $[252]$ and could be used in the same way as the calculation for $\alpha = 2$. 

127
dimension of the inclusive jet function is known to three loops \cite{66, 277}. We present the relevant expressions in App. B.4.

This leaves the soft function \(S_G(z_{\text{cut}})\) and the collinear-soft function \(S_C(z_{\text{cut}}e_2^{(2)})\) to be determined. Their one-loop expressions are easily calculable, and we present the results in App. B.3 and App. B.5. To determine their two-loop non-cusp anomalous dimensions, we exploit the renormalization group consistency of the factorization theorem. The sum of the anomalous dimensions must vanish at each order:

\[
0 = \gamma_H + \gamma_S + 2\gamma_f + 2\gamma_{SC},
\]

where \(\gamma_F\) denotes the anomalous dimension of function \(F\) in the factorization theorem, and we have used the symmetry of the left and right hemispheres of the event. Therefore, only one unknown anomalous dimension remains, which we take to be \(\gamma_S\).

### 3.5.1.1 Two-Loop Soft Function

To calculate the two-loop non-cusp anomalous dimension \(\gamma_S\) we need to calculate the soft function \(S_G(z_{\text{cut}})\) with two real emissions. The two-loop expression for the soft function is

\[
S_G(z_{\text{cut}})|_{a_2} = \int [d^dk_1][d^dk_2]_{+}|\mathcal{M}(k_1, k_2)|^2\Theta_{\text{SD}}.
\]

Here, \([d^dk_1]_{+}\) is the positive-energy on-shell phase space measure in \(d = 4 - 2\epsilon\) dimensions:

\[
[d^dk_1]_{+} = \frac{d^dk_1}{(2\pi)^d}2\pi\delta(k_1^2)\Theta(k_1^0),
\]
and $|\mathcal{M}(k_1, k_2)|^2$ is the squared matrix element for two soft emissions from a $q\bar{q}$ dipole. The explicit expression for $|\mathcal{M}(k_1, k_2)|^2$ can be found in Ref. [112]. $\Theta_{SD}$ is the phase space constraint imposed by the soft drop groomer. Recall that, for consistency with the assumed hierarchy $e_2^{(a)} \ll z_{\text{cut}}$, soft modes must fail soft drop.

If the particles in the hemispheres are reclustering using the Cambridge/Aachen algorithm, $\Theta_{SD}$ can be written as

$$
\Theta_{SD} = \Theta(-\eta_1 \eta_2) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) 
$$

$$
+ \Theta(\eta_1 \eta_2) \left[ \Theta(\theta_{1,J} - \theta_{12}) \Theta(\theta_{2,J} - \theta_{12}) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 - k_2^0 \right) \right. 
$$

$$
+ \left. [1 - \Theta(\theta_{1,J} - \theta_{12}) \Theta(\theta_{2,J} - \theta_{12})] \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) \right].
$$

The first line of Eq. (3.5.7) corresponds to particles 1 and 2 lying in different hemispheres (opposite rapidity with respect to the $q\bar{q}$ dipole), and so each particle individually must fail soft drop. $Q$ is the center of mass energy and so $Q/2$ is the energy in one hemisphere. The second and third lines correspond to the configuration where both particles lie in the same hemisphere. $\theta_{12}$ is the angle between the particles and $\theta_{i,J}$ (for $i = 1, 2$) is the angle particle $i$ makes with that hemisphere’s axis. If $\theta_{12}$ is less than both $\theta_{1,J}$ and $\theta_{2,J}$ then, according to the Cambridge/Aachen algorithm, the soft particles are clustered first. Therefore, the sum of the energies of particles 1 and 2 must fail soft drop. If instead one of the particles is closer to the jet axis, then they are clustered separately and must individually fail soft drop.

To proceed, we separate the squared matrix element into Abelian and non-Abelian
pieces, according to their color coefficient. At this order, the squared matrix element takes the form

\[ |\mathcal{M}(k_1, k_2)|^2 = |\mathcal{M}_{\text{n-A}}(k_1, k_2)|^2 + \frac{1}{2!} |\mathcal{M}(k_1)|^2 |\mathcal{M}(k_2)|^2 , \quad (3.5.8) \]

Here, “n-A” denotes the non-Abelian component of the squared matrix element, which includes the \( C_F C_A \) and \( C_F n_f T_R \) color channels. The Abelian contribution is just the symmetrized product of the one-loop result, with a color factor of \( C_F^2 \). We will consider these two pieces separately, starting with the non-Abelian term.

**Non-Abelian Clustering Effects**

Note that except for the effects from Cambridge/Aachen clustering, soft drop is just imposing a soft energy veto on each hemisphere. The two-loop soft function with a soft energy veto was calculated in Ref. [309]. That calculation showed that the two-loop Abelian piece (proportional to \( C_F^2 \)) to the energy vetoed soft function satisfies non-Abelian exponentiation. The two-loop non-cusp anomalous dimension for a hemisphere energy vetoed soft function is then purely non-Abelian and was extracted in Ref. [127]. The non-Abelian part of the soft function with an energy veto at two-loops is

\[ S_{\text{veto}}|_{\text{n-A}, \alpha_s^2} = \int [d^d k_1]_+ [d^d k_2]_+ |\mathcal{M}_{\text{n-A}}(k_1, k_2)|^2 \Theta_{\text{veto}} . \quad (3.5.9) \]

The phase space cut \( \Theta_{\text{veto}} \) is

\[ \Theta_{\text{veto}} = \Theta \left( \Lambda - k_1^0 - k_2^0 \right) , \]
where $\Lambda$ is the veto scale. We can then write the two-loop soft function for soft drop as

$$S_G(z_{\text{cut}})|_{\Lambda, \alpha_s^2} = S_{\text{veto}}|_{\Lambda, \alpha_s^2} + \int [d^4k_1] + [d^4k_2] + |M_{n, A}(k_1, k_2)|^2 [\Theta_{SD} - \Theta_{\text{veto}}] ,$$

(3.5.10)

where the veto scale is set to $\Lambda = z_{\text{cut}} Q/2$. The difference between the soft drop and energy veto phase space constraints is purely a clustering effect, given by

$$\Theta_{SD} - \Theta_{\text{veto}} = \{ \Theta(\eta_1 \eta_2) [1 - \Theta(\theta_1 J - \theta_1 J) \Theta(\theta_2 J - \theta_1 J)] + \Theta(-\eta_1 \eta_2) \} \times \Theta\left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta\left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) \Theta\left( k_1^0 + k_2^0 - z_{\text{cut}} \frac{Q}{2} \right).$$

Eq. (3.5.10) enables us to calculate much more simply the two-loop non-cusp anomalous dimension of the soft function. The anomalous dimension can then be written as

$$\gamma_S = \gamma_{\text{veto}} + \gamma_{C/A} .$$

(3.5.12)

$\gamma_{\text{veto}}$ is the two-loop non-cusp anomalous dimension of $S_{\text{veto}}$ extracted in Ref. [127]:

$$\gamma_{\text{veto}}^2 = \left( \frac{\alpha_s}{4\pi} \right)^2 C_F \left[ \left( \frac{1616}{27} - 56 \zeta_3 \right) C_A - \frac{448}{27} n_f T_R - \frac{2\pi^2}{3} \beta_0 \right] ,$$

(3.5.13)

where $\beta_0$ is the one-loop $\beta$-function coefficient:

$$\beta_0 = \frac{11}{3} C_A - \frac{4}{3} n_f T_R .$$

(3.5.14)

Then, we only need to determine the contribution to the anomalous dimension from residual Cambridge/Aachen clustering effects, $\gamma_{C/A}$. 

131
The non-Abelian clustering effects are contained in

$$S_G(z_{\text{cut}})_{n-A, \alpha_s^2}^{C/A} = \int [d^d k_1] + [d^d k_2] + \left| \mathcal{M}_{n-A}(k_1, k_2) \right|^2 \left[ \Theta_{SD} - \Theta_{\text{veto}} \right].$$

(3.5.15)

The squared non-Abelian matrix element does not have collinear singularities when the angle of the particles from the jet axis is strongly ordered. Therefore, in this integral there is only a collinear divergence when the two emissions become collinear to the jet axis in a non-strongly ordered way. The coefficient of this divergence is proportional to the correction to the two-loop anomalous dimension due to clustering effects in the non-Abelian color channel. The divergence can be extracted with the standard plus-function prescription and the correction to the anomalous dimension can be found. While we were unable to find an analytic expression, its approximate numerical value is\(^4\)

$$S_G(z_{\text{cut}})_{n-A, \alpha_s^2}^{C/A} = \left( \frac{\alpha_s}{4\pi} \right)^2 C_F \left[ -9.31 C_A - 14.04 n_f T_R \right] \left( \frac{4 \mu^2}{z_{\text{cut}}^2 Q^2} \right)^{2 \epsilon} \frac{1}{4 \epsilon} + \mathcal{O}(\epsilon^0).$$

(3.5.16)

The contribution to the anomalous dimension is then

$$\gamma_{C/A}^{n-A, \alpha_s^2} = \left( \frac{\alpha_s}{4\pi} \right)^2 C_F \left[ -9.31 C_A - 14.04 n_f T_R \right].$$

(3.5.17)

\(^4\)This anomalous dimension does not seem to be a linear combination of the usual transcendental numbers appearing in other two-loop anomalous dimensions.
Abelian Clustering Effects

The Abelian contribution can be calculated similarly. However, unlike the non-Abelian contribution, the exponentiation of the one-loop result will describe at least some of the two-loop Abelian piece. If the square of the one-loop result does not account for all of the two-loop result, then non-Abelian exponentiation breaks down. This does not mean that exponentiation breaks down or that the cross section cannot be resummed, just that the anomalous dimension of the purely Abelian piece will need to be corrected at every logarithmic order. So, for the two-loop non-cusp anomalous dimension, we need to determine the part of the soft function that is not accounted for by non-Abelian exponentiation.

To do this, we start from the full expression for the Abelian term at two-loops:

$$S_G(z_{\text{cut}})_{A,\alpha^2} = \frac{1}{2!} \int [d^dk_1] + [d^dk_2] + |\mathcal{M}(k_1)|^2 |\mathcal{M}(k_2)|^2 \Theta_{SD} . \tag{3.5.18}$$

We then add and subtract the one-loop phase space constraints:

$$S_G(z_{\text{cut}})_{A,\alpha^2} = \frac{1}{2!} \int [d^dk_1] + [d^dk_2] + |\mathcal{M}(k_1)|^2 |\mathcal{M}(k_2)|^2 \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right)$$

$$+ \frac{1}{2!} \int [d^dk_1] + [d^dk_2] + |\mathcal{M}(k_1)|^2 |\mathcal{M}(k_2)|^2 \left[ \Theta_{SD} - \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) \right] . \tag{3.5.19}$$

The difference between the phase space constraints is a clustering effect, given by

$$\Theta_{SD} - \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) \tag{3.5.20}$$

$$= -\Theta(\eta_1 \eta_2) \Theta(\theta_{1J} - \theta_{12}) \Theta(\theta_{2J} - \theta_{12})$$

133
\[
\Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) \Theta \left( k_1^0 + k_2^0 - z_{\text{cut}} \frac{Q}{2} \right) .
\]

As with the non-Abelian term, this phase space constraint completely removes all soft divergences and the strongly-ordered collinear limit. The remaining divergence can be isolated by standard plus-function techniques. For the two-loop Abelian Cambridge/Aachen clustering term, we find the numerical result

\[
S_G(z_{\text{cut}})_{A, \alpha_s}^{C/A} = \left( \frac{\alpha_s}{4\pi} \right)^2 34.01 C_F^2 \left( \frac{4\mu^2}{z_{\text{cut}}^2 Q^2} \right)^{2\epsilon} \frac{1}{4\epsilon} + \mathcal{O}(\epsilon^0), \tag{3.5.21}
\]

for the second integral in Eq. (3.5.19). The contribution to the anomalous dimension is then

\[
\gamma_{C/A}^{A, \alpha_s^2} = \left( \frac{\alpha_s}{4\pi} \right)^2 34.01 C_F^2 . \tag{3.5.22}
\]

### 3.5.1.2 Two-Loop Anomalous Dimension

Combining Eqs. (3.5.13), (3.5.17) and (3.5.22), the total two-loop non-cusp anomalous dimension for the soft function is

\[
\gamma_S^{\alpha_s^2} = \left( \frac{\alpha_s}{4\pi} \right)^2 C_F \left[ 34.01 C_F + \left( \frac{1616}{27} - 56\zeta_3 - 9.31 \right) C_A - \left( \frac{448}{27} + 14.04 \right) n_f T_R - \frac{2\pi^2}{3} \beta_0 \right] . \tag{3.5.23}
\]

The two-loop non-cusp anomalous dimension for the collinear-soft function is found by consistency using Eq. (3.5.4). Note that this anomalous dimension has no \( \log z_{\text{cut}} \) terms. Therefore, the anomalous dimensions of no functions in the factorization theorem have \( \log z_{\text{cut}} \) dependence. This is a consequence of the fact that each function of our factorization theorem in Eq. (3.3.10) depends on a single infrared scale, al-
Following NNLL resummation of all logarithms of $z_{\text{cut}}$ and $e_2^{(2)}$ alike. As we discuss in Secs. 3.3.3 and 3.5.2, this result relies on the choice of Cambridge/Aachen reclustering in the soft drop algorithm.

We can verify this result by comparing the resummed distribution, truncated at $O(\alpha_s^2)$, with the singular region of the full QCD result, computed to the same fixed order. For the full QCD result, we have implemented soft drop into EVENT2 [113], a Monte Carlo code that generates fixed-order results up to $O(\alpha_s^2)$ in $e^+e^-$ collisions. Our specific implementation is as follows. We generate $e^+e^-$ collisions at 1 TeV center of mass energy and identify event hemispheres with the exclusive $k_T$ algorithm [110].

We then recluster each hemisphere using the Cambridge/Aachen algorithm and apply soft drop with $\beta = 0$. On each of the soft-drop groomed hemispheres, we then measure the energy correlation function $e_2^{(2)}$ and record the larger of the two values, which we denote by $e_{2,H}^{(2)}$ and refer to as the heavy groomed mass. This is simply related to the cross section of our factorization theorem:

$$\frac{d\sigma}{de_2^{(2),H}} = \frac{d^2\sigma}{de_{2,L}^{(2)} de_{2,R}^{(2)}} \frac{d^2\sigma}{de_{2,L}^{(2)} de_{2,R}^{(2)}} \left[ \Theta \left( e_{2,L}^{(2)} - e_{2,R}^{(2)} \right) \delta \left( e_{2,H}^{(2)} - e_{2,L}^{(2)} \right) + (L \leftrightarrow R) \right].$$

(3.5.24)

In Fig. 3.4, we compare EVENT2 results to the prediction of the factorized expression at NNLL expanded to $O(\alpha_s^2)$. For soft drop with $\beta = 0$, soft logarithms are removed, which means that at $O(\alpha_s^2)$, the cross section has the schematic form

$$e_{2,H}^{(2)} \frac{d\sigma}{de_2^{(2),H}} \sim \alpha_s^2 C_0 \log e_{2,H}^{(2)} + \alpha_s^2 C_1,$$

(3.5.25)
Figure 3.4: Verification of our factorization theorem at $\mathcal{O}(\alpha_s^2)$ for soft-drop grooming with $z_{\text{cut}} = 0.001$ and $\beta = 0$. Solid curves are numerical results from EVENT2, and dashed curves are $\mathcal{O}(\alpha_s^2)$ terms in our NNLL distribution, plotted in the three color channels $C_F^2$, $C_F C_A$, and $C_F n_f T_R$. (a) shows a direct comparison and (b) the difference.

where $C_0$ and $C_1$ are constants. We plot the cross section separated into the three color channels ($C_F^2$, $C_F C_A$, and $C_F n_f T_R$). We set $z_{\text{cut}} = 0.001$ to suppress power corrections of $z_{\text{cut}}$. Excellent agreement between our factorization theorem and EVENT2 is observed at small $e_2^{(2)}$, demonstrating that we have captured all singular terms of the full QCD result in our factorization theorem to $\mathcal{O}(\alpha_s^2)$.

### 3.5.2 Reclustering with anti-$k_T$

It is illuminating to study the clustering effects in the soft function in more detail.

In this section, we re-calculate the clustering effects with the anti-$k_T$ algorithm, instead of the standard Cambridge/Aachen algorithm. We find that the clustering effects with the anti-$k_T$ algorithm are intimately related to the corresponding effects
calculated in jet veto calculations. This can be understood relatively simply by re-expressing the clustering conditions in a form analogous to the clustering metric of the longitudinally-invariant $k_T$ algorithm.

To calculate the two-loop soft function for soft drop defined with anti-$k_T$ re-clustering, we only need to calculate the clustering effects unique to this algorithm. We will denote the phase space constraints for the anti-$k_T$ re-clustering as $\Theta^{\text{ak}_T}$, but we will not explicitly present them here. The two-loop soft function is

$$S^{\text{ak}_T}(z_{\text{cut}}) = \left. \Theta^{\text{ak}_T} \right|_{\alpha_s^2} = \int [d^d k_1] + [d^d k_2] + |\mathcal{M}(k_1, k_2)|^2 \Theta_{\text{veto}} + \int [d^d k_1] + [d^d k_2] + |\mathcal{M}(k_1, k_2)|^2 \left[ \Theta^{\text{ak}_T} - \Theta_{\text{veto}} \right].$$

The relevant phase space constraints can be written as

$$\Theta^{\text{ak}_T} - \Theta_{\text{veto}} = \Theta(\eta_1 \eta_2) \left[ 1 - \Theta \left( \max[k_1^0, k_2^0] \theta_{1J} - \frac{Q}{2} \theta_{12} \right) \Theta \left( \max[k_1^0, k_2^0] \theta_{2J} - \frac{Q}{2} \theta_{12} \right) \right]$$

$$\times \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_1^0 \right) \Theta \left( z_{\text{cut}} \frac{Q}{2} - k_2^0 \right) \Theta \left( k_1^0 + k_2^0 - z_{\text{cut}} \frac{Q}{2} \right).$$

With this, we can calculate the divergent part of the two-loop soft function from clustering effects and extract the anomalous dimension. As with Cambridge/Aachen, we can write the two-loop non-cusp anomalous dimension as

$$\gamma_S^{\text{ak}_T} = \gamma_{\text{veto}} + \gamma_{\text{ak}_T},$$

where $\gamma_{\text{ak}_T}$ is the part of the anomalous dimension purely from clustering effects. We find

$$\gamma_{\text{ak}_T} = -8 \left( \frac{\alpha_s}{4\pi} \right)^2 C_F \left\{ \left[ \left( \frac{131}{9} - \frac{4}{3} n_f \pi^2 - \frac{44}{3} \log 2 \right) C_A + \left( -\frac{46}{9} + \frac{16}{3} \log 2 \right) n_f T_R \right] \log z_{\text{cut}} \right\}.$$
\[
+ \left( \frac{-269}{6} + \frac{7}{2} \zeta_3 + \frac{274}{9} \log 2 + \frac{11\pi^2}{9} + \frac{44}{3} \log^2 2 \right) C_A \\
+ \left( \frac{53}{3} - \frac{4\pi^2}{9} - \frac{116}{9} \log 2 - \frac{16}{3} \log^2 2 \right) n_f T_R \right) .
\] (3.5.29)

This anomalous dimension is fascinating. First, note that there is no \( C_F^2 \) term, implying that non-Abelian exponentiation holds for anti-\( k_T \) re-clustering, in contrast to what we found for the Cambridge/Aachen algorithm. That is, all logarithms at \( \mathcal{O}(\alpha_s^2) \) with color factor \( C_F^2 \) are accounted for by exponentiating the one-loop result. This is to be expected: Since the anti-\( k_T \) algorithm clusters soft gluons (with energy fractions of order \( z_{\text{cut}} \)) one-by-one with the hard jet core unless two soft gluons have angular separation \( \Delta R \lesssim z_{\text{cut}} \), clustering effects are merely a power correction for Abelian gluons.

Also, unlike the case for Cambridge/Aachen re-clustering, there is explicit log \( z_{\text{cut}} \) dependence in the anomalous dimension of Eq. (3.5.29). This shows that we do not resum logarithms of \( z_{\text{cut}} \) to full NNLL accuracy when anti-\( k_T \) clustering is used in soft drop. The coefficient of the log \( z_{\text{cut}} \) term is identical to the coefficient of the logarithm of the jet radius \( R \) found from clustering effects in jet veto calculations [46, 49, 65, 67, 302]. This connection between soft drop and jet veto calculations can be made clearer by a simple rewriting of the clustering metric.

The \( k_T \) class of clustering metrics for \( e^+e^- \) collisions can be written as

\[
d_{ij} = \min \left[ E_i^{2p}, E_j^{2p} \right] \theta_{ij}^2,
\] (3.5.30)

138
for particles $i$ and $j$, with $p$ an integer that defines the jet algorithm. In the soft function, soft particles are either clustered with each other or with the jet axis. For $\beta = 0$, these soft particles have characteristic energy fraction $z_{\text{cut}} \ll 1$. In terms of energy fractions, the clustering metric of two soft particles is

$$d_{ij} = \min \left[ z^p_i, z^p_j \right] Q^{2p} \theta^2_{ij} \sim z^p_{\text{cut}} Q^{2p} \theta^2_{ij},$$

(3.5.31)

and for a soft particle $i$ with the jet axis it is

$$d_i = \min[1, z^p_{\text{cut}}] Q^{2p} \theta^2_i,$$

(3.5.32)

where $\theta_i$ is the angle between particle $i$ and the jet axis.

Consider $p < 0$. In this case, the two soft gluons are (parametrically) clustered together when

$$z^p_{\text{cut}} \theta_{ij} < \min[\theta_i, \theta_j],$$

(3.5.33)

or equivalently, when

$$\theta_{ij} < z^{\lvert p \rvert}_{\text{cut}} \min[\theta_i, \theta_j].$$

(3.5.34)

The effective clustering metric in this case is then

$$d_{ij}^{\text{eff}} = \min \left[ z^p_i, z^p_j \right] \frac{\theta^2_{ij}}{z^{|p|}_{\text{cut}} \min[\theta^2_i, \theta^2_j]}, \quad d_i^{\text{eff}} = z^p_i.$$

(3.5.35)

With $p = -1$, this is the clustering metric for the inclusive anti-$k_T$ algorithm with effective jet radius $R = z_{\text{cut}} \ll 1$. There will now be logarithms of the jet radius that arise. The log $z_{\text{cut}}$ term in the anomalous dimension has the identical coefficient as
the log $R$ term in jet veto calculations because $z_{\text{cut}}$ and $R$ act as the angular scale for collinear splittings in the respective soft functions.

In summary, while we could use anti-$k_T$ to recluster the jet for soft drop grooming, we could not resum all large logarithms to the same precision without a different factorization theorem. Therefore, reclustering in soft drop with the Cambridge/Aachen algorithm is preferred from a theory perspective.

### 3.5.3 NNLL for $\alpha = 2$, $\beta \geq 0$

For soft drop with angular exponent $\beta > 0$, we cannot recycle results from the literature to reach NNLL precision. Instead, a completely new two-loop calculation of either the soft or collinear-soft function is needed. But without such a calculation, we can perform NNLL resummation for particular values of $\beta > 0$, using numerical simulations to estimate the ingredients we lack. We will demonstrate this explicitly in the case of $\beta = 1$, and the result will allow us to study features of NNLL distributions for energy correlation functions with less aggressive grooming.

The same method we used to validate anomalous dimensions for $\beta = 0$ can be used to extract the anomalous dimension for $\beta > 0$. This method relies on the fact that all ingredients necessary for NNLL resummation with $\alpha = 2$, $\beta > 0$ are known except the two-loop non-cusp anomalous dimensions of the soft and collinear-soft functions. As mentioned above, renormalization group invariance determines one of these, say $\gamma^{(1)}_{SC}$,
in terms of the other anomalous dimensions. So only one unknown, \( \gamma_S^{(1)} \), remains and we can extract it at fixed order.

To do this for a given \( \beta > 0 \), we can use EVENT2 to obtain numerical results at \( \mathcal{O}(\alpha_s^2) \) for the groomed \( e_{2,H}^{(2)} \) distribution with several moderately small values of \( z_{\text{cut}} \). From each of these distributions, we can subtract the known terms, which we get by expanding the NNLL distribution to fixed order. This leaves a term proportional to the unknown \( \gamma_S^{(1)} \), as well as power corrections suppressed by \( e_{2,H}^{(2)} \) or \( z_{\text{cut}} \). By computing the distribution down to very small \( e_{2,H}^{(2)} \), we can ignore the \( e_{2,H}^{(2)} \) power corrections. Reducing power corrections from \( z_{\text{cut}} \) is limited by the numerical precision of EVENT2 because our factorization theorem only applies for \( e_{2,H}^{(2)} \ll z_{\text{cut}} \). Instead, we can fit the \( z_{\text{cut}} \) power corrections to linear combinations of \( z_{\text{cut}} \log^n(z_{\text{cut}}) \log^m(e_{2,H}^{(2)}) \).

At \( \mathcal{O}(\alpha_s^2) \) it is appropriate to use \( 0 \leq n + m \leq 3 \), though in practice we found terms with \( m \geq 2 \) to be difficult to fit. With the non-negligible power corrections thus removed, we can then extract the remaining anomalous dimension.

While the procedure outlined above is straightforward, an explicit calculation of \( \gamma_S^{(1)} \) or \( \gamma_{SC}^{(1)} \) for \( \beta > 0 \) is of course desirable. On practical time scales, numerical extractions are limited to rough approximations, due to inadequate numerical precision in the deep infrared. Nevertheless, an estimate is sufficient for our purposes here, which are to demonstrate the advantages of resumming jet substructure observables to NNLL, and to examine various levels of grooming. Thus, we will test the above procedure on \( \beta = 0 \), and learn about the associated uncertainties by comparing with
Figure 3.5: Demonstration of non-cusp anomalous dimension extraction in EVENT2. (a) Solid curves are numerical results from EVENT2 at $\mathcal{O}(\alpha_s^2)$ with $\beta = 0$ and $z_{\text{cut}} = 0.1$. Dashed curves are $\mathcal{O}(\alpha_s^2)$ terms in NNLL distribution, without the term proportional to $\gamma_S^{(1)}$. Discrepancy results from $z_{\text{cut}}$ power corrections in solid curves and missing $\gamma_S^{(1)}$ in dashed curves. Subtracting $z_{\text{cut}} \log^n(z_{\text{cut}}) \log(e_{2,H}^{(2)})$ power corrections from dashed curves, we extract the remaining offsets. (b) As $z_{\text{cut}} \to 0$, remaining offsets allow extraction of $\gamma_S^{(1)}$ in rough agreement with Eq. (3.5.23).

our direct calculation, Eq. (3.5.23). Then we will move to $\beta = 1$, and extract $\gamma_S^{(1)}$ in that case.

In Fig. 3.5a we show numerical results at $\mathcal{O}(\alpha_s^2)$ from EVENT2 with $\beta = 0$ and $z_{\text{cut}} = 0.1$. Also shown is the NNLL distribution, expanded to fixed order, but without the $\gamma_S^{(1)}$ term. The discrepancy between the curves is thus due to the missing $\gamma_S^{(1)}$ term and $z_{\text{cut}}$ power corrections. Using several distributions like this one, with values of $z_{\text{cut}}$ between $10^{-4}$ and $10^{-1}$, we fit the $z_{\text{cut}}$ power corrections. Fig. 3.5b shows the remaining offsets between our analytical curves and the results of EVENT2, after $z_{\text{cut}} \log^n(z_{\text{cut}}) \log(e_{2,H}^{(2)})$ power corrections have been subtracted. On each point in this

142
<table>
<thead>
<tr>
<th>Soft Drop $\gamma_S^{(1)}$</th>
<th>$C_F$</th>
<th>$C_A$</th>
<th>$n_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = 0$ extraction</td>
<td>$28 \pm 1.5$</td>
<td>$-40 \pm 1$</td>
<td>$-23 \pm 3$</td>
</tr>
<tr>
<td>$\beta = 0$ calculation</td>
<td>$34.01$</td>
<td>$-40.90$</td>
<td>$-21.86$</td>
</tr>
<tr>
<td>$\beta = 1$ extraction</td>
<td>$6 \pm 12$</td>
<td>$-9.5 \pm 2$</td>
<td>$-8 \pm 7$</td>
</tr>
</tbody>
</table>

Table 3.2: Extraction of two-loop non-cusp anomalous dimension $\gamma_S^{(1)}$ of wide-angle soft function in the three different color channels. For $\beta = 0$ comparison with our direct calculation is possible. See text for discussion of uncertainties.

plot, the error bar represents the standard deviation in EVENT2 output, across $e_{2,H}^{(2)}$ bins. The offset that remains as $z_{\text{cut}} \to 0$ is the $\gamma_S^{(1)}$ we would extract using this method. One can see from Fig. 3.5b that agreement with our analytical calculation, Eq. (3.5.23), is quite good, with some discrepancy in the $C_F$ channel.

Table 3.2 lists the numerical results of $\gamma_S^{(1)}$ using this method.\(^5\) The uncertainties quoted for the $\beta = 0$ extraction in the table come from the standard deviation in EVENT2 output across $e_{2,H}^{(2)}$ bins, which introduces an error in the identification of constant offsets. These should be compared with our direct calculation in the second line of the table. The discrepancy in the $C_F$ channel gives us a sense of additional numerical uncertainties, which are significant. Similar disagreements have been encountered before, e.g. in Ref. [130], in the context of $C_F$ channel extractions from EVENT2, and to resolve it might require significantly longer run times.

\(^5\)In carrying out the procedure just described, we tuned EVENT2 parameters to favor the infrared. In particular, we use of order 1 trillion events, with $\text{CUTOFF} = 10^{-15}$ and phase-space sampling exponents $\text{NPOW1} = \text{NPOW2} = 5$. This procedure corresponded to centuries of CPU time.
As stated above, a rough estimate of $\gamma_S^{(1)}$ for $\beta > 0$ is sufficient for our purposes, so we have applied the method described above to the case of $\beta = 1$. See the third line of Table 3.2 for the results of the extraction. Uncertainties quoted in this line of the table have two sources: (i) variance in EVENT2 output, and (ii) additional numerical precision issues, which we took to be the difference (both absolute and relative) between extraction and direct calculation in the $\beta = 0$ test. In each color channel, we took the maximum of these uncertainties and inflated it by a factor of 2.

The estimate in Table 3.2 allows us to study NNLL distributions of $\ell_{2,H}^{(2)}$ groomed with $\beta = 1$. In the resulting distributions, the uncertainties associated with the imperfect extraction are relatively small; e.g. see Fig. 3.6b below. Still, a direct calculation of either $\gamma_S^{(1)}$ or $\gamma_{S_C}^{(1)}$ for $\beta > 0$ would of course be preferred, but we leave this to future work.

### 3.6 Matching NNLL to Fixed Order

Using the results calculated in the previous sections, here we match our resummed differential cross section for soft-drop groomed energy correlation functions to fixed-order for hemisphere jets produced in $e^+e^-$ collisions. We first match resummed results at NLL and NNLL to $O(\alpha_s)$ and $O(\alpha_s^2)$, respectively, using EVENT2 and demonstrate that theoretical uncertainties are greatly reduced at NNLL. We then compare several Monte Carlo parton shower simulations to our matched NNLL results. We
compare both parton and hadron level Monte Carlo to our perturbative analytic results, and include a simple model of hadronization in our calculation. We leave a detailed understanding and justification of incorporating hadronization into the resummed and matched cross section to future work.

### 3.6.1 Matching Resummation to Fixed-Order

With the explicitly calculated and extracted two-loop non-cusp anomalous dimensions of the soft function in the soft drop factorization theorem Eq. (3.3.10), we are able to resum the differential cross section through NNLL accuracy in the region where $e_2^{(2)} \ll z_{\text{cut}} \ll 1$. Anomalous dimensions of all functions are collected in the appendices and we present the explicit form of the resummed cross section in App. B.6.

This resummed cross section is only valid in the region where $e_2^{(2)} \ll z_{\text{cut}} \ll 1$, and will not provide an accurate description of the cross section outside this region. To accurately describe the cross section throughout the full phase space requires matching the resummed result to fixed-order.

While there are many ways to do this at various levels of sophistication, we choose to use simple additive matching. That is, we construct matched distributions according to

$$\frac{d\sigma_{\text{match}}}{de_2^{(2)}} = \frac{d\sigma_{\text{resum}}}{de_2^{(2)}} + \frac{d\sigma_{\text{FO}}}{de_2^{(2)}} - \frac{d\sigma_{\text{resum,FO}}}{de_2^{(2)}}.$$  \hspace{1cm} (3.6.1)

Here, $d\sigma_{\text{resum}}$ is the resummed cross section, calculated to the appropriate logarith-
mic accuracy. $d\sigma_{\text{FO}}$ is the fixed-order differential cross section calculated to a particular order in $\alpha_s$. $d\sigma_{\text{resum,FO}}$ is the resummed cross section truncated at the same accuracy as the fixed-order cross section. In the infrared phase space region, this term will exactly cancel the singularities in the fixed-order cross section, only leaving the resummed cross section plus power corrections. In the hard phase space region, this term cancels the resummed cross section, up to higher orders in $\alpha_s$.

The logarithmic accuracy of the resummed cross section was defined in Sec. 3.5, and here we specify the fixed orders that we use in the matching procedure. We additively match the analytic NLL distributions to $O(\alpha_s)$ fixed order results, which include one real emission from the $q\bar{q}$ dipole. We match NNLL distributions to $O(\alpha_s^2)$ results, which include up to two real emissions. EVENT2 is able to generate $e^+e^-$ collisions through $O(\alpha_s^2)$, except for the two-loop virtual contribution. The two-loop virtual term only contributes at $e_2^{(2)} = 0$, so our differential distributions are unaffected by this omission.

In Fig. 3.6, we plot the resummed and matched differential cross sections for the larger $e_2^{(2)}$ of the two hemispheres at NLL and NNLL with various levels of soft drop grooming. Here, we consider dijet production in $e^+e^-$ collisions at 1 TeV center-of-mass energy and identify hemispheres with the exclusive $k_T$ algorithm [110]. The parameters of soft drop are $z_{\text{cut}} = 0.1$ and we show both $\beta = 0$ and $\beta = 1$. We also show the ungroomed heavy hemisphere $e_2^{(2)}$ distribution. In these plots, we include estimates of theoretical uncertainties represented by the lighter bands about the central
Figure 3.6: (a) NLL matched distributions for heavy hemisphere $e_2^{(2)}$ in $e^+e^-$ collisions with soft drop grooming $z_{\text{cut}} = 0.1$ and $\beta = 0$, $\beta = 1$, and without soft drop. Estimates of theoretical uncertainties are represented by the shaded bands. (b) The corresponding matched distributions at NNLL. For soft drop with $\beta = 1$, the dotted lines represent the extent of the theoretical uncertainties when the variation of the two-loop non-cusp anomalous dimension is included. Note the significant reduction in uncertainties at NNLL.

curve. While more sophisticated methods for estimating uncertainties exist, we simply vary the natural scales that appear in the functions of the factorization theorem up and down by a factor of two. We then take the envelope of these scale variations as an estimate of theoretical uncertainties. This simple prescription is sufficient for our main purpose in showing uncertainty bands: to demonstrate the reduction in theoretical uncertainty in moving from NLL to NNLL.

Included in these uncertainty estimates is a variation in our treatment of the Landau pole of the strong coupling $\alpha_s$. For scales $\mu > 1$ GeV, $\alpha_s$ is evaluated according to its perturbative running. For $\mu < 1$ GeV, we freeze $\alpha_s$ to its value at the scale
\( \mu = 1 \text{ GeV} \). This is not intended to be a model for hadronization or non-perturbative physics, but is just intended to maintain finite cross section predictions at small \( \epsilon^2 \) values. To estimate the sensitivity of our results to the scale at which we freeze the coupling, we vary this 1 GeV scale by a factor of two, and include the effect in the uncertainty bands of Fig. 3.6 as well.

Finally, we have shown the uncertainty bands around the \( \beta = 1 \) curves at NNLL with and without the uncertainty in our estimate of the two-loop non-cusp anomalous dimension of the soft function. One can see from the figure that this imperfect extraction has only a relatively small effect on the overall uncertainty at this order.

Importantly, we allow the normalization of the cross section to change under these scale variations. That is, the curves in Fig. 3.6 are constructed according to Eq. (3.6.1). The normalization of each distribution displayed is meaningful, since we resum all large logs in both the shape and the normalization. While the central value curves don’t change much in going from NLL to NNLL, the uncertainties are dramatically reduced, and this is partly due to the increased accuracy in the normalization.

### 3.6.2 Comparison to Monte Carlo

In this section, we compare our NNLL resummed and matched soft-drop groomed \( \epsilon^2 \) distributions to the output of several standard Monte Carlo simulations. We generate \( e^+e^- \rightarrow \text{dijets} \) events at 1 TeV center-of-mass collision energy with HERWIG++
Figure 3.7: Comparison between soft-drop groomed $e_2^{(2)}$ distributions with $z_{cut} = 0.1$ and $\beta = 0$ for NNLL, parton-level, and hadron-level Monte Carlo. All curves integrate to the same value over the range $e_2^{(2)} \in [0.01, 1]$. 

149
2.7.1 [41, 72], Pythia 8.210 [2, 295], and VINCIA 1.2.02 [198, 199, 214, 247]. While the Herwig++ and Pythia events are showered from the leading order process \( e^+e^- \rightarrow q\bar{q} \), we consider VINCIA with and without fixed-order matching included. The matched VINCIA results are accurate effectively through \( \mathcal{O}(\alpha_s^2) \). For the most direct comparison of the simulations to our NNLL matched results, we run \( \alpha_s \) at two loops in all Monte Carlos (in the CMW scheme [116, 162]) and we fix \( \alpha_s(m_Z) = 0.118 \), which is the same value used in our analytic calculations. We include Monte Carlo events both at partonic level and after hadronization. These events are then clustered into hemispheres using the exclusive \( k_T \) algorithm [110] using FastJet 3.1.3 [99]. The soft drop grooming and subsequent measurement of the two-point energy correlation functions of these \( e^+e^- \) events is implemented in FastJet with custom code.

We compare the Monte Carlo distributions to our NNLL resummed and matched calculations in Figs. 3.7 and 3.8. In these plots, all distributions integrate to the same value over the range \( e^{(2)}_2 \in [0.01, 1] \). In Fig. 3.7, we compare the soft-drop groomed \( e^{(2)}_2 \) distributions with \( z_{\text{cut}} = 0.1, \beta = 0 \). Good agreement between the Monte Carlos and our analytic calculation is observed, with (not surprisingly) the matched Monte Carlo agreeing the best. These distributions also show that parton- and hadron-level Monte Carlos are essentially identical for \( e^{(2)}_2 \gtrsim 0.001 \). In Fig. 3.8, we compare the soft-drop groomed \( e^{(2)}_2 \) distributions with \( z_{\text{cut}} = 0.1, \beta = 1 \). Again, good agreement between the Monte Carlos and our matched NNLL result is observed, with the parton- and hadron-level Monte Carlos nearly identical for \( e^{(2)}_2 \gtrsim 0.005 \). The uncer-
Figure 3.8: Comparison between soft-drop groomed $e_2^{(2)}$ distributions with $z_{\text{cut}} = 0.1$ and $\beta = 1$ for NNLL, parton-level, and hadron-level Monte Carlo. All curves integrate to the same value over the range $e_2^{(2)} \in [0.01, 1]$. The uncertainty band for NNLL includes the variation of the two-loop non-cusp anomalous dimension.
Figure 3.9: Direct comparison of hadron-level output from HERWIG++, PYTHIA, and VINCIA already shown in Figs. 3.7 and 3.8. Soft drop is performed with $z_{\text{cut}} = 0.1$ and both $\beta = 0$ (left) and $\beta = 1$ (right). Curves are displayed as relative differences between Monte Carlo output and our matched NNLL predictions, with theoretical uncertainties shown as a shaded band.

tainty bands for the analytic curve includes the uncertainty in the two-loop non-cusp anomalous dimension.

As a more direct comparison of the Monte Carlos, Fig. 3.9 displays the relative difference between each of the hadron-level Monte Carlos and our matched NNLL predictions. Again, soft drop is performed with $z_{\text{cut}} = 0.1$, and both $\beta = 0$ and $\beta = 1$ are shown. All the Monte Carlo curves lie within our shaded band of theoretical uncertainty, but discrepancies between the different simulations are visible.

One striking feature in these plots, especially for $\beta = 0$, is the presence of additional structure in the hadron-level Monte Carlo distributions at small $e_2^{(2)}$. It is clear that this feature is due to non-perturbative physics, and so is therefore not included in our NNLL calculation. Nevertheless, we can include a simple model of hadroniza-
tion into our calculation to see if this structure is easily explained.

For additive IRC safe observables, like thrust or jet mass, it can be shown from general principles that hadronization corrections can be incorporated in perturbative distributions by convolution with a model shape function [240, 241]. In general, the energy correlation functions are additive observables, so we should be able to use shape functions to model hadronization corrections. However, once soft drop is applied on the jet, emissions in the jet may or may not contribute to the energy correlation functions, so the observable is no longer strictly additive. We leave a more careful study of whether shape functions can be used to model hadronization effects in groomed observables to future work. Here we convolve our matched results with a simple shape function to see if qualitative agreement with the Monte Carlos can be achieved.

Because shape functions describe non-perturbative physics, they only have support for energies comparable to $\Lambda_{QCD}$. The shape function we choose is the parametrization suggested by Ref. [301]:

$$F_{\text{shape}}(\epsilon) = \frac{4\epsilon}{\Omega^2} e^{-2\epsilon/\Omega}. \quad (3.6.2)$$

This is normalized

$$\int_0^\infty d\epsilon F_{\text{shape}}(\epsilon) = 1, \quad (3.6.3)$$

and has first moment equal to $\Omega$. As discussed in Sec. 3.4.3, of all modes present in our factorization theorem, the collinear-soft mode has the lowest virtuality, so it will
Figure 3.10: Perturbative NNLL results for soft-drop groomed $e_2^{(2)}$ with $z_{cut} = 0.1$ and $\beta = 0$ (left) and $\beta = 1$ (right), compared to analytic results that include the shape function of Eq. (3.6.2) for modeling hadronization, and compared to hadron-level Monte Carlo. The parameter $\Omega = 1$ GeV. Note that, qualitatively, the shape function produces a hadronization-bump similar to those seen in the Monte Carlos.

have the largest sensitivity to non-perturbative physics. We thus convolve our perturbative distribution with the shape function, assuming non-perturbative effects are primarily associated with the collinear-soft mode. That is, we include hadronization corrections in the soft drop groomed $e_2^{(2)}$ distribution according to

$$
\frac{d\sigma_{\text{had}}}{de_2^{(\alpha)}} = \int de \frac{d\sigma_{\text{pert}}}{de_2^{(\alpha)}} \left( e_2^{(\alpha)} - \left( \frac{e}{z_{\text{cut}} Q} \right)^{\alpha - 1} \right) F_{\text{shape}}(e),
$$

(3.6.4)

where the argument of the perturbative distribution is shifted by the virtuality of the collinear-soft mode, Eq. (3.4.7).

In Fig. 3.10, we compare the matched NNLL distribution of $e_2^{(2)}$ with and without convolution with the shape function of Eq. (3.6.2), in which we set $\Omega = 1$ GeV to
be comparable to the scale of hadron masses. We show this comparison for soft drop grooming with $\beta = 0$ and $\beta = 1$. The peak at small $e_2^{(2)}$ for $\beta = 0$ agrees qualitatively with the structure of the hadronized Monte Carlo distributions. Similarly, the shape at small $e_2^{(2)}$ for $\beta = 1$ agrees with the simulations as well. This suggests that there might exist a shape function for describing hadronization effects in groomed jet observables, though we leave a detailed discussion and justification for such a model to future work.

3.7 Matching to Fixed Order in $pp \to Z + j$

In this section, we present predictions for soft-drop groomed $e_2^{(\alpha)}$ distributions as measured on the jet in $pp \to Z + j$ events at the LHC. The definitions of soft drop and energy correlation functions appropriate for jets in $pp$ collisions are given in Sec. 3.2. As with jets from $e^+e^-$ collisions, we match our NNLL resummed distribution to fixed-order results that include relative $\mathcal{O}(\alpha_s^2)$ corrections to the Born process.

There are two complications we must deal with. First, at $pp$ collisions, the jets will be both quark and gluon initiated. Second, because we only measure the observable within the jet and do not constrain radiation throughout the rest of the event, the simple hard-soft-jet factorization that we employed for $e^+e^- \to$ hemisphere jets will not apply here. Nevertheless, while the normalization of the jet-observable distribution will thus be complicated and sensitive to multiple scales, the shape of the dis-
tribution will still be controlled exclusively by collinear physics. To address these complications, we first show how soft-drop groomed quark and gluon jets can be unambiguously defined order-by-order in perturbation theory. Then we discuss how the normalization of the distribution can be obtained by matching to full QCD at fixed order. The discussion will focus on the $Z + j$ sample for concreteness, but these ideas apply equally well to any process with hard jets at a hadron collider.

### 3.7.1 Resummed Cross Section in $pp \to Z + j$

We define our observable on soft-drop groomed jets in $pp \to Z + j$ events in the following way. First, we cluster the final state according to a jet algorithm with some jet radius $R \sim 1$. Of the jets with pseudorapidity $|\eta_I| < \eta_{\text{max}}$, we then identify the jet with the largest transverse momentum $p_{T,I}$ and require that $p_{T,I} > p_{T}^{\text{min}}$. We groom this jet with soft drop and measure $e_2^{(\alpha)}$ according to the definitions given in Sec. 3.2 for jets in $pp$ collisions. In this procedure, we remain inclusive over all other hadronic activity in the final state: we only care about the hardest jet.

For this process, the relevant factorization formula is

$$\frac{d\sigma_{\text{resum}}}{de_2^{(\alpha)}} = \sum_{k=q,\bar{q},g} D_k(p_{T}^{\text{min}}, \eta_{\text{max}}, z_{\text{cut}}, R)S_{C,k}(z_{\text{cut}}e_2^{(\alpha)}) \otimes J_k(e_2^{(\alpha)}). \tag{3.7.1}$$

This cross section reproduces and resums the terms in the full QCD cross section that are singular in the limit $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$. We now explain the components of this formula in detail.
In Eq. (3.7.1), \( S_{C,k}(z_{\text{cut}}e_2^{(\alpha)}) \) and \( J_k(e_2^{(\alpha)}) \) are the collinear-soft and jet functions for the measurement of soft-drop groomed \( e_2^{(\alpha)} \) that, by collinear factorization, are identical to the functions defined in \( e^+e^- \) collisions. Unlike in \( e^+e^- \) collisions, however, these functions also have a label \( k \) corresponding to the flavor of the jet, and a sum over the possible QCD parton flavors \( k \) is included. The symbol \( \otimes \) denotes convolution in \( e_2^{(\alpha)} \) between the collinear-soft and jet functions.

\( D_k \) is a matching coefficient that can be extracted from fixed-order calculations, and it sets the normalization and relative contributions from the different jet flavors. In addition to the dependence explicitly shown, \( D_k \) also depends implicitly on parton distributions, as different initial states produce different flavors of final state jets.

Unlike the case in \( e^+e^- \) collisions, where the jet energy was (almost exactly) half the center-of-mass energy, due to the non-trivial parton distributions, the distribution of the jet \( p_T \) has a finite width and depends on the cut, \( p_T^{\text{min}} \). For a true precision prediction, we would compute the matching coefficient \( D_k \) as a function of \( p_{T,J} \) and include an integral in Eq. (3.7.1) convolving the jet and collinear-soft functions with \( D_k(p_{T,J}) \). An approach to doing this in a semi-automatic manner was discussed recently in Refs. [61, 171]. But, for simplicity we instead employ the following approximation: we evaluate the jet and collinear-soft functions at \( \langle p_{T,J} \rangle \), the average \( p_{T,J} \).

The average jet transverse momentum \( \langle p_{T,J} \rangle \) can be estimated by using the fact that
the cross section for a jet with transverse momentum $p_{T,J}$ takes the power-law form:

$$\frac{1}{\sigma} \frac{d\sigma}{dp_{T,J}} \simeq \frac{n-1}{p_{T,J}^{min}} (\frac{p_{T,J}^{min}}{p_{T,J}})^n \Theta(p_{T,J} - p_{T,J}^{min}).$$  \hfill (3.7.2)

This distribution is normalized and the mean value of $p_{T,J}$ is

$$\bar{p}_{T,J} = \frac{n-1}{n-2} p_{T,J}^{min}. \hfill (3.7.3)$$

The typical exponent is $n \sim 5$, and we take $n = 5$ in our numerical computations.

The full cross section for soft-dropped $e_2^{(\alpha)}$ (including power corrections) can be expressed as

$$\frac{d\sigma}{de_2^{(\alpha)}} = \sum_{k=q,g} D_k S_{C,k} \otimes J_k + \frac{d\sigma_{pc}}{de_2^{(\alpha)}}. \hfill (3.7.4)$$

Here, the right-most term includes all power corrections suppressed by $e_2^{(\alpha)}$ or $z_{cut}$.

The functions $S_{C,k}$ and $J_k$ should be evaluated at $\bar{p}_{T,J}$ but we have suppressed their arguments for brevity. We will use this form of the cross section to define the matching coefficient $D_k$ at fixed-order. For NNLL resummation, the relative $O(\alpha_s)$ corrections to $D_k$ are required.

First, at leading order in $\alpha_s$, Eq. (3.7.4) becomes

$$\frac{d\sigma^{(0)}}{de_2^{(\alpha)}} = \sum_{k=q,g} D_k^{(0)} \delta(e_2^{(\alpha)}), \hfill (3.7.5)$$

where the superscript $(0)$ denotes the leading order in $\alpha_s$. Here, we have used $J_k^{(0)} = S_{C,k}^{(0)} = \delta(e_2^{(\alpha)})$. Also, since a jet has only one constituent at this order, the distribution has no support away from $e_2^{(\alpha)} = 0$ and there are no partons to soft drop; therefore, there are no $e_2^{(\alpha)}$ or $z_{cut}$ power corrections at this order. Integrating over all $e_2^{(\alpha)}$,
we are left with the Born-level cross section for the $k$ flavor channel $\sigma_k^{(0)}$, so that

$$D_k^{(0)} = \sigma_k^{(0)}.$$

At the next-to-leading order in $\alpha_s$, the extraction of $D_k$ requires separating the jets by flavor. Since $D_k$ is defined in each flavor channel, we need to determine the flavor of the hardest jet in each $pp \rightarrow Z + j$ event included in our sample. Ordinarily, any definition of jet flavor based on the constituents of the jet is infrared-unsafe and ill-defined at leading power, because soft wide-angle emissions into a jet can change its flavor.\(^6\) Soft drop eliminates this problem at leading power in $e_2^{(\alpha)}$ and $z_{\text{cut}}$ by removing soft wide-angle radiation from the jet. This allows for an infrared and collinear safe definition of jet flavor at leading power in $e_2^{(\alpha)}$ and $z_{\text{cut}}$. We define the jet flavor $f_J$ as the flavor sum of the constituents of the groomed jet:

$$f_J = \sum_{i \in J_g} f_i,$$

where $f_q = 1$, $f_{\bar{q}} = -1$ and $f_g = 0$. The subscript on $J_g$ means that one only sums over the jet constituents that remain after grooming with soft drop. If $f_J = \pm 1$, then the jet is quark-type, while if $f_J = 0$, it is gluon-type. With this jet flavor identification, we are able to determine the total fixed-order cross section for each jet flavor channel in $pp \rightarrow Z + j$. We will denote the next-to-leading order term in the cross section for a jet of flavor $k$ as $\sigma_k^{(1)}$, defined according to the phase space cuts described at the beginning of this section.

\(^6\)However, one infrared and collinear safe definition of jet flavor was presented in Ref. [48].
Then, at next-to-leading order in the $k$ flavor channel, Eq. (3.7.4) becomes

$$\frac{d\sigma_k^{(1)}}{d\alpha_2} = D_k^{(0)} \left[ S_{C,k}^{(1)} + J_k^{(1)} \right] + D_k^{(1)} \delta(e_2^{(1)}) \cdot \frac{d\sigma_k^{(1)}}{d\alpha_2}. \quad (3.7.8)$$

Here, $S_{C,k}^{(1)}$ and $J_k^{(1)}$ are the collinear-soft and jet functions at $O(\alpha_s)$. Using $D_k^{(0)} = \sigma_k^{(0)}$, we can integrate over $e_2^{(\alpha)}$ to find

$$D_k^{(1)} = \sigma_k^{(1)} - \sigma_k^{(0)} \int_0^1 d\alpha_2 \left[ S_{C,k}^{(1)} + J_k^{(1)} \right] - \sigma_{k,pc}^{(1)}. \quad (3.7.9)$$

We computed $\sigma_k^{(1)}$ using MCFM [103, 104] with settings detailed in the next section.

We computed the power corrections according to

$$\sigma_{k,pc}^{(1)} \equiv \int d\alpha_2 \left[ \frac{d\sigma_k^{(1)}}{d\alpha_2} - \sigma_k^{(0)} \left( J_k^{(1)} + S_{C,k}^{(1)} \right) \right]. \quad (3.7.10)$$

For the first term in the integrand, we use a numerical distribution obtained with MCFM. Since we do not have access to this distribution at arbitrarily small values of $e_2^{(\alpha)}$, the integral in Eq. (3.7.10) extends from $e_2^{(\alpha)} = 10^{-5}$ to 1. This approximation is sufficient for power corrections suppressed by $e_2^{(\alpha)}$, and the effect of dropping the $z_{\text{cut}} \delta(e_2^{(\alpha)})$ term from the integral is negligible in comparison to the scale uncertainties shown in the next section.

This completes our extraction of the matching coefficient $D_k$ through relative $O(\alpha_s)$. With it, the resummed cross section of Eq. (3.7.1) is complete and ready to be matched to relative $O(\alpha_s^2)$ fixed-order results.
3.7.2 Matching Resummation to Fixed-Order

With the resummed differential cross section for soft-drop groomed $e_2^{(\alpha)}$ defined in Eq. (3.7.1), we next match to fixed order for $pp \rightarrow Z + j$. Our matching procedure will be identical to the procedure we used for $e^+e^-$ collisions; we add the difference between the exact fixed order and the expansion of the resummed distribution to fixed order:

$$
\frac{d\sigma_{\text{match}}}{de_2^{(\alpha)}} = \frac{d\sigma_{\text{resum}}}{de_2^{(\alpha)}} + \frac{d\sigma_{\text{FO}}}{de_2^{(\alpha)}} - \frac{d\sigma_{\text{resum,FO}}}{de_2^{(\alpha)}}. \quad (3.7.11)
$$

We match the analytic NLL resummed distributions to fixed-order results that include the relative $O(\alpha_s)$ corrections to the Born process for $pp \rightarrow Z + j$. We match NNLL distributions to fixed-order results including relative $O(\alpha_s^2)$ corrections and up to 3 partons in the jet.

We use MCFM v. 6.8 [103, 104] to generate the fixed-order cross sections for soft-drop groomed $e_2^{(\alpha)}$ in $pp \rightarrow Z + j$ events. Currently, MCFM can only generate fixed-order corrections at $O(\alpha_s)$ relative to a Born-level process, and so we will have to use some properties of the observable to be able to calculate to relative $O(\alpha_s^2)$ accuracy. For $e_2^{(\alpha)} > 0$, as we did in $e^+e^-$ collisions, we can ignore the purely two-loop virtual contribution to $pp \rightarrow Z + j$, as it has no effect on the differential distribution away from $e_2^{(\alpha)} = 0$. MCFM can generate both inclusive $pp \rightarrow Z + j$ and $pp \rightarrow Z + 2j$ processes through relative $O(\alpha_s)$ accuracy. Therefore, we can use $pp \rightarrow Z + 2j$ at relative $O(\alpha_s)$ in MCFM to calculate the relative $O(\alpha_s^2)$ distribution for $pp \rightarrow Z + j$,
in the region where $e_2^{(\alpha)} > 0$.

In practice, this procedure requires some care. To define the cross section for $pp \to Z + 2j$ in MCFM, we must set a minimum $p_T$ for the two jets as identified by MCFM. This is set by the parameter $\text{ptjet\_min}$ within MCFM. To compute the fixed-order cross section correctly for $e_2^{(\alpha)}$ as measured on the soft-drop groomed jet in $pp \to Z + j$ events, $\text{ptjet\_min}$ should be set to 0; this would of course produce infinity because $pp \to Z + 2j$ lacks the virtual corrections of $pp \to Z + j$. To regulate this divergence, we set $\text{ptjet\_min} = 1$ GeV and have verified that for jets with $p_{T,J} > 500$ GeV, this choice has a negligible effect on the differential cross section of $e_2^{(\alpha)}$ until deep in the infrared region, well beyond the point where resummation dominates. Additionally, we have verified that the distribution of $e_2^{(\alpha)}$ as measured in $pp \to Z + j$ at relative $\mathcal{O}(\alpha_s^2)$ is identical to that measured in $pp \to Z + 2j$ at Born level with $\text{ptjet\_min} = 1$ GeV, up to differences deep in the infrared. Using this procedure, we are therefore able to match to relative $\mathcal{O}(\alpha_s^2)$ with MCFM.

We generate $pp \to Z + j$ events through relative $\mathcal{O}(\alpha_s^2)$ accuracy at the 13 TeV LHC using MSTW 2008 NLO parton distribution functions [265]. We require that the $p_T$ of the $Z$ boson is greater than 300 GeV and the absolute value of its pseudorapidity is less than 2.5. Jets are clustered with the anti-$k_T$ algorithm with radius $R = 0.8$. We study the hardest jet in these events that satisfies $p_{T,J} > 500$ GeV and $|\eta_J| < 2.5$. On these identified jets, we then soft-drop groom and measure $e_2^{(\alpha)}$ using custom code. This is an exceptionally computationally demanding procedure at relative $\mathcal{O}(\alpha_s^2)$, due
Figure 3.11: NLL matched (left) and NNLL matched (right) distributions for hardest jet $e_2^{(2)}$ in $pp \rightarrow Z + j$ events with soft drop grooming $z_{\text{cut}} = 0.1$ and $\beta = 0$ and $\beta = 1$. Estimates of theoretical uncertainties are represented by the shaded bands. For soft drop with $\beta = 1$, the dotted lines represent the extent of the theoretical uncertainties when the variation of the two-loop non-cusp anomalous dimension is included. The distributions in the two upper figures are normalized to the total cross section (in femtobarns), while in the bottom figures, the distributions integrate to the same value over the range $e_2^{(2)} \in [0.001, 0.1]$. Note the reduction in uncertainties as one moves from NLL to NNLL, and also as one considers the normalized distribution.
to the complicated phase space of real emissions and the small width of the bins required to calculate the $e_2^{(a)}$ distribution. This precision jet substructure study is only possible because of the development of highly efficient methods for generating fixed-order corrections.

In Fig. 3.11 we plot matched distributions for soft-drop $e_2^{(2)}$ with $z_{\text{cut}} = 0.1$ and both $\beta = 0$ and $\beta = 1$ at NLL and NNLL. Here, we show both the distributions normalized to the total cross section and normalized over the range $e_2^{(2)} \in [0.001, 0.1]$. The shaded bands represent estimates of theoretical uncertainties due to residual infrared scale sensitivity.\footnote{The relatively large size of the uncertainty bands for $e_2^{(2)} \gtrsim 0.1$ is an artifact of our simplistic additive matching. Additionally, due to the large $K$ factor, the absolute scale of the matched NNLL distribution in Fig. 3.11b is roughly twice as large as the matched NLL distribution in Fig. 3.11a.} We show these bands mainly to allow comparison of the uncertainty remaining at different levels of formal precision. For the collinear-soft and jet functions in the resummed cross section, we vary the low scales by a factor of two. To estimate the scale dependence of the matching coefficient $D_k$ in the resummed cross section is more complicated, and we discuss this in detail in App. B.7. To estimate scale uncertainties in the fixed-order cross section, we vary the factorization and renormalization scales in MCFM by a factor of 2 about 500 GeV $\simeq p_{T,J}$. We then take the envelope of all of these scale variations to produce the shaded bands in Fig. 3.11. For $\beta = 1$ at NNLL, we have also explicitly shown the additional uncertainty due to the two-loop non-cusp anomalous dimension of the collinear-soft function. In go-
ing from NLL to NNLL accuracy, the relative size of the scale uncertainty bands decreases by about a factor of 2 or 3 for both choices of normalization of the distributions. However, normalizing the distributions over the range $e_2^{(2)} \in [0.001, 0.1]$ dramatically reduces residual scale uncertainties; at NNLL, these normalized distributions have residual scale uncertainties at the 10\% level and smaller.

### 3.7.3 Comparison to Monte Carlo

We now compare our NNLL resummed and matched calculation of soft-drop groomed $e_2^{(2)}$ distributions to Monte Carlo simulations. We generate $pp \rightarrow Z + j$ events at the 13 TeV LHC with HERWIG++ 2.7.1 and PYTHIA 8.210. To improve statistics somewhat, we have turned off $Z/\gamma$ interference in the Monte Carlos. The $Z$ boson is forced to decay to electrons, and we require that the invariant mass of the electrons is within 10 GeV of the mass of the $Z$ boson. We then require that the identified $Z$ boson has $p_{TZ} > 300$ GeV and $|\eta_Z| < 2.5$. Jets are clustered with FASTJET 3.1.3 using the anti-$k_T$ algorithm with radius $R = 0.8$ and we identify the hardest jet in the event with $p_{T,J} > 500$ GeV and $|\eta_J| < 2.5$. We then soft-drop groom this jet and measure $e_2^{(2)}$. Both soft drop and the energy correlation functions are implemented using FASTJET contrib v. 1.019 [99].

We have generated two samples from both HERWIG++ and PYTHIA to study the effect of hadronization and underlying event. One sample is purely parton level: both
Figure 3.12: Comparison between soft-drop groomed $e_2^{(2)}$ distributions with $z_{\text{cut}} = 0.1$ and $\beta = 0$ (top) and $\beta = 1$ (bottom) for matched and normalized NNLL, parton-level, and hadron-level Monte Carlo. All curves integrate to the same value over the range $e_2^{(2)} \in [0.001, 0.1]$. The uncertainty band for soft drop with $\beta = 1$ at NNLL includes the variation of the two-loop non-cusp anomalous dimension.
Figure 3.13: Direct comparison of hadron-level output from HERWIG++ and PYTHIA already shown in Fig. 3.12. Soft drop is performed with $z_{\text{cut}} = 0.1$ and both $\beta = 0$ (left) and $\beta = 1$ (right). Curves are displayed as relative differences between Monte Carlo output and our matched NNLL predictions, with theoretical uncertainties shown as a shaded band.

hadronization and underlying event have been turned off and the other sample is the Monte Carlos run in their default settings, up to the settings of the $Z$ boson mentioned earlier. The distributions of $e^{(2)}_2$ measured on soft-drop groomed jets with $z_{\text{cut}} = 0.1$ and both $\beta = 0, 1$ are illustrated in Fig. 3.12. Here, we compare our matched and normalized NNLL calculation to both the parton-level and hadron-level plus underlying event Monte Carlos. To normalize the Monte Carlo distributions, all curves integrate to the same value on the range $e^{(2)}_2 \in [0.001, 0.1]$.

As a more direct comparison of the Monte Carlos, Fig. 3.13 displays the relative difference between each of the hadron-level Monte Carlos and our matched NNLL predictions, with our estimates of theoretical uncertainty shown as shaded bands. Again, soft drop is performed with $z_{\text{cut}} = 0.1$, and both $\beta = 0$ and $\beta = 1$ are shown. Dis-
crepancies between the Monte Carlo results and our predictions are present but not large.

As observed with jets in $e^+e^-$ collisions, there is good agreement between our precision calculation and the Monte Carlos over a wide dynamic range. Importantly, this measurement of the soft-drop groomed $e_2^{(2)}$ is very different from the case in $e^+e^-$. In $e^+e^-$ collisions we calculated the heavy groomed and ungroomed jet masses. By measuring the heavier of the two jet masses, both masses have to be small, and the observable is global. For $pp \to Z + j$ events, we want to make no restrictions on the out-of-jet radiation. Thus although the soft drop jet mass is still free of non-global contributions, the ungroomed mass will not be. That is, we do not have control over all the large logarithms of ungroomed jet mass in $pp \to Z + j$ events, and thus cannot predict them using our factorized expression, although other approaches are possible.\footnote{Calculations of the ungroomed jet mass in $Z + j$ events have been done, with varying approaches to handing the non-global contribution \cite{128, 149, 224}.}

For this reason, we only show distributions of soft-drop groomed $e_2^{(2)}$ measurements in $pp \to Z + j$ events.

Fig. 3.12 also illustrates that soft drop grooming eliminates sensitivity to both hadronization and underlying event until deep in the infrared. The parton-level and hadron-level distributions for each Monte Carlo agree almost perfectly until below about $e_2^{(2)} \lesssim 10^{-3}$. That hadronization effects are small is expected from our $e^+e^-$ analysis, but this also demonstrates that underlying event effects are negligible.

\footnote{Calculations of the ungroomed jet mass in $Z + j$ events have been done, with varying approaches to handing the non-global contribution \cite{128, 149, 224}.}
similar observation was made in Ref. [248], though at a much higher jet $p_T$ ($p_T > 3$ TeV). As in $e^+e^-$ collisions, we expect that the hadronization effects that are observed in the Monte Carlo can be explained by a shape function, though we leave this to future work.

That the shape of the resummed distribution is both completely determined by collinear dynamics and is insensitive to underlying event suggests that by grooming jets with soft drop, we are able to completely isolate factorization-violating effects into an overall normalization. Therefore, we conjecture that the shape of the leading-power distribution of soft-drop groomed observables as measured in hadron collision events completely factorizes, just like the $p_T$ spectrum in Drell-Yan events [142]. We leave a proof of this conjecture to future work.\(^9\)

### 3.8 Conclusions

In this paper, we presented the first calculation for an observable measured exclusively on the constituents of a jet to NNLL accuracy and matched to fixed-order results at $O(\alpha_s^2)$ relative to the Born process. The ability to do this calculation required grooming the jet with the soft drop algorithm, which eliminates the complications

\(^9\)Due to the presence of the complicated object $D_k(p_T^{\text{min}}, \eta_{\text{max}}, z_{\text{cut}}, R)$, Eq. (3.7.1) is not strictly a factorization theorem. It may not be possible to factorize $D_k$ to all orders due to the presence of so-called Glauber modes [142] in the cross section. While it is beyond the scope of this paper, recent work suggests that Glaubers can be included into the cross section directly [290], and our numerical work indicates that the effect may be absorbable into the normalization.
due to non-global logarithms that afflict ungroomed jet measurements. The soft drop groomer also significantly reduces nonperturbative effects from hadronization and underlying event, rendering the perturbative calculation of energy correlation functions accurate over several decades. The insensitivity of soft-drop groomed jet observables to underlying event suggests that the normalized cross section fully factorizes in hadronic scattering events.

To complete the resummed calculation to NNLL accuracy required determining the two-loop non-cusp anomalous dimension for the soft function for which all emissions are removed. For $\beta = 0$, we were able to use results from the literature to extract the non-cusp anomalous dimension, up to calculable clustering effects. While not used for results in this paper, the clustering effects when using the anti-$k_T$ algorithm with soft drop are closely related to similar effects found in jet veto calculations. For soft drop angular exponent $\beta > 0$, we demonstrated a numerical procedure for determining the anomalous dimension using EVENT2. This was sufficient to approximate the non-cusp anomalous dimension, but a full calculation of the two-loop soft function for soft drop with $\beta \geq 0$ is desired.

With a complete calculation of the two-loop soft function, including constants, we would be one step closer to resumming to next-to-next-to-next-to-leading logarithmic accuracy (N$^3$LL). Up to the unknown four-loop cusp anomalous dimension (whose effects have been shown to be small [21, 69, 216]), the only other piece to get to N$^3$LL would be the three-loop non-cusp anomalous dimension of the soft-dropped soft func-
tion. Without an explicit three-loop calculation, this anomalous dimension could in principle be estimated using a technique similar to what we used at two loops, using a fixed-order code like EERAD3 [196]. If this is possible, then resummation to this accuracy would potentially reduce residual scale uncertainties to the percent-level, assuming a scaling of uncertainties like observed in going from NLL to NNLL.

For our complete predictions, it was vital to match our resummed calculations to high precision fixed-order distributions. Fixed-order calculations have been traditionally used for observables that are inclusive over soft and collinear radiation, like total cross sections or $p_T$ spectra. The generation of fixed-order differential distributions for the plots in this paper required CPU-centuries, which we attained only by running on thousands of cores. For calculations of more complicated jet observables, precise fixed order computations are likely infeasible with presently available tools. As jet substructure pushes to higher precision, it will be necessary to have fixed-order calculations that more efficiently sample the infrared regions of phase space.

The calculations in this paper represent a new frontier of precision QCD. While jet substructure techniques have been used for some time in experimental analyses at the LHC, they are just now approaching the level of theoretical precision that can be meaningfully compared to data. By soft-drop grooming jets, we greatly reduce the theoretical challenges, enabling the calculation of a wide range of jet substructure observables to full NNLL accuracy.

171
Optimization of Calculable Quark/Gluon Discriminants
4.1 Introduction

The fantastic jet reconstruction performance of ATLAS and CMS [10, 233]—along with increasingly sophisticated tools to predict jet properties from first principles [146, 148, 152, 187, 250, 254, 287, 310]—has led to significant advances in the field of jet substructure [22, 24, 32, 33]. A key goal in jet substructure is to robustly discriminate quark-initiated jets from gluon-initiated jets [36, 84, 155, 191, 192, 205, 237, 253, 255], with many applications to new physics searches at the Large Hadron Collider (LHC) (see e.g. [85, 139, 178]). In the eikonal limit, quarks and gluons differ only by their respective color charges, $C_F = 4/3$ versus $C_A = 3$, such that gluon jets emit more soft gluon radiation than quark jets. At this order, the difference between quark and gluon radiation patterns is controlled entirely by the Casimir ratio $C_A/C_F = 9/4$, which drives (and limits) the expected separation power between quark and gluon jets.

One of the most powerful quark/gluon discriminants is hadron multiplicity, or its charged-particle-only variant, track multiplicity $n_{tr}$ [16, 19, 140, 191, 192, 255]. This is an effective discriminant because the average track multiplicity within quark and gluon jets scales approximately as (see e.g. [89, 90])

$$\frac{\langle n_{tr}\rangle_q}{\langle n_{tr}\rangle_g} \approx \frac{C_A}{C_F}.$$  \hspace{1cm} \hspace{1cm} (4.1.1)

Since multiplicity is not infrared and collinear (IRC) safe, though, it is difficult to predict its discrimination performance from first principles.\footnote{It is possible to calculate the evolution with energy of the multiplicity moments; see, e.g.,} On the other hand, IRC-
safe observables like jet mass and jet width are analytically tractable \([47, 170, 252]\), but they exhibit worse quark/gluon performance than multiplicity. The reason is that these discriminants are dominated by a single emission at leading-logarithmic (LL) accuracy, giving rise to Casimir scaling of the quark/gluon discrimination power,

\[
(\text{gluon mistag rate}) \simeq (\text{quark efficiency})^{C_A/C_F},
\]

and therefore relatively weak separation between quark and gluon jets. This Casimir scaling behavior holds for any observable with a Sudakov form factor at LL accuracy, including a wide range of IRC-safe additive observables \([253]\). While one can try to interpolate between the IRC-unsafe and IRC-safe regimes using generalized angularities \([255]\), track multiplicity remains one of the best performing—yet analytically puzzling—quark/gluon discriminants.

In this paper, we introduce a new class of “counting observables” that are IRC safe, yet yield comparable quark/gluon performance to track multiplicity. Unlike additive observables, which are only sensitive to a single emission at LL order, these counting observables are directly sensitive to multiple emissions at LL, allowing them to exceed the performance estimate in Eq. (4.1.2). Crucially, the quark/gluon performance of counting observables still depends on the color factors \(C_A\) and \(C_F\), but instead of being described by Sudakov form factors, these observables are described by Poisson distributions; this allows their discrimination power to improve as more emissions are

Ref. [166] for a review.
included. These counting observables not only clarify the underlying reason why track multiplicity performs so well, but they also demonstrate the new kinds of analytic structures possible from IRC-safe but non-additive observables.\(^2\)

The counting observables we study are based on an iterated variant of soft drop declustering [248]. As a grooming procedure, soft drop starts at the trunk of an angular-ordered clustering tree [163, 312] and sequentially removes soft branches with small momentum fraction \(z_{ij}\) until a hard branching is found. At a step in the clustering tree where branches \(i\) and \(j\) split, the splitting is retained in the groomed jet if the momentum fraction satisfies

\[
z_{ij} > z_{\text{cut}} \left( \frac{\theta_{ij}}{R_0} \right)^\beta,
\]

where \(\theta_{ij}\) is an appropriately defined relative angle between branches \(i\) and \(j\), and \(R_0\) is the jet radius. For appropriate choices of the soft drop parameters \(z_{\text{cut}}\) and \(\beta\), observables defined on the groomed jet are automatically infrared (but not necessarily collinear) safe. While the original soft drop procedure terminates once it finds a hard \(1 \to 2\) splitting, the iterated variant we employ in this paper continues, following the hardest branch (the “trunk”) through multiple levels until an angular cutoff scale \(\theta_{\text{cut}}\) is reached.

The simplest counting observable we can define using iterated soft drop (ISD) is

\(^2\)An alternative counting method was proposed in Ref. [84], which considers associated subjets outside of the jet boundary. Additionally, there has been interest in understanding the scaling of the cross section at high jet multiplicity [17, 39, 79, 197]. Here, we focus on counting subjets within the jet of interest.
just the total number of emissions from the trunk of the clustering tree that ISD
records. In particular, this includes all emissions \( n \in [1, n_{\text{max}}] \) that satisfy the soft
drop condition and lie outside the \( \theta_{\text{cut}} \) cone. We call this observable “soft drop multi-

\[
n_{\text{SD}}(z_{\text{cut}}, \beta, \theta_{\text{cut}}) = \sum_{n} 1, \quad (4.1.4)
\]

which depends on the choice of ISD parameters. It is complementary to the “soft drop
level” observable \( L_{\text{SD}}(\beta) \) introduced in Ref. [18], which also iteratively applies the soft
drop condition, but changes the \( z_{\text{cut}} \) scale. As long as \( z_{\text{cut}} > 0 \), soft drop multiplicity
is infrared safe.

With \( \theta_{\text{cut}} > 0 \) or \( \beta < 0 \), \( n_{\text{SD}} \) is collinear safe as well, so we can use analytic re-
summation tools to predict its discrimination power. We do this to resum large loga-
rithms of \( z_{\text{cut}} \) and \( \theta_{\text{cut}} \), which are of soft and collinear origin, respectively, and which
lead to a double-logarithmic observable. The analysis at LL order is straightforward,
yielding a Poisson distribution whose average value is set by the phase space “area” of
counted emissions. This leads to quark/gluon discrimination power which approaches
that of track multiplicity, particularly in the case of \( \beta = -1 \). Moving from LL to
next-to-leading-logarithmic (NLL) order, one finds a slight decrease in discrimination
power, due in part to the jet-flavor mixing that appears at this accuracy. We imple-
ment the NLL calculation through a set of evolution equations that have a similar
form to parton evolution.
With $\theta_{\text{cut}} = 0$ and $\beta \geq 0$, the soft drop multiplicity $n_{\text{SD}}$ is no longer collinear safe, so we cannot predict its absolute discrimination power. That said, for the special case of $\beta = 0$ (which was initially introduced as the modified mass drop tagger [147, 148]), we can use renormalization group (RG) techniques to predict the evolution of its discrimination power. When $\beta = 0$, soft drop multiplicity has purely collinear divergences, which can be absorbed into a generalized fragmentation function (GFF) that depends on the RG scale $\mu$ [165]. After extracting this GFF at low scales (either from LHC data\(^3\) or parton shower simulations), one can use a perturbative DGLAP-like evolution equation to predict the discrimination power achievable at higher scales. Intriguingly, in the limit of pure Yang-Mills, one can show that at lowest order, the soft drop multiplicity asymptotes to a true Poisson distribution at large values of $\mu$, such that it behaves like an idealized counting observable (albeit in a theory with only gluons).

The remainder of this paper is organized as follows. In Sec. 4.2, we define the ISD procedure, introduce soft drop multiplicity, and take a first look at its distribution using parton shower generators. In Sec. 4.3, we perform an LL analysis, focusing on the contrast between soft drop multiplicity’s Poisson behavior and the more familiar Sudakov-peak behavior of additive observables. We extend our analytic calculations

\(^3\)Just as for parton distribution functions and ordinary fragmentation functions, extracting GFFs involves matching to fixed-order calculations, as described in Ref. [165]. These fixed-order calculations involve a mixture of quark and gluon final-state partons, so multiple event samples with different quark/gluon fractions are required to disentangle the contributions from quark and gluon GFFs.
to NLL order in Sec. 4.4 and compare our analytic distributions to those obtained from various parton showers. We consider the collinear-un-safe case of $\theta_{\text{cut}} = 0$ and $\beta = 0$ in Sec. 4.5, deriving the corresponding RG evolution equations and presenting numerical results based on parton shower inputs. We present our conclusions in Sec. 4.6.

In an appendix, we demonstrate that our analytical tools can also be used to study more general ISD observables, in particular the weighted multiplicity $\sum_n (z_n)^{\kappa}$ which weights each counted emission according to its momentum fraction $z_n$. Soft drop multiplicity is a special case ($\kappa = 0$) of this more general observable, and the one most useful for quark/gluon discrimination.

### 4.2 Counting Observables with Soft Drop

#### 4.2.1 Iterated Soft Drop

Our counting observables are defined using an iterated variant of the soft drop declustering algorithm. We briefly review soft drop here for convenience and to establish conventions.

The soft drop grooming procedure can be applied to any jet found using a standard jet algorithm of characteristic radius $R_0$. After reclustering the jet using the Cambridge/Aachen (C/A) algorithm [163, 312], soft drop involves sequentially undoing
the cluster history to remove wide-angle soft radiation and identify hard 2-prong substructure. For each C/A branching into subjets \( i \) and \( j \), there are quantities \( z_{ij} \) and \( \theta_{ij} \), which are defined differently for different collider environments:

\[
e^+ e^- \text{ collisions: } z_{ij} = \frac{\min(E_i, E_j)}{E_i + E_j}, \quad \theta_{ij} = \text{angle between } i, j, \quad (4.2.1)
\]

\[
pp \text{ collisions: } z_{ij} = \frac{\min(p_{T_i}, p_{T_j})}{p_{T_i} + p_{T_j}}, \quad \theta_{ij} = \Delta R_{ij}, \quad (4.2.2)
\]

where \( \Delta R \) represents distance in the rapidity-azimuth plane. The soft drop grooming algorithm can be summarized as follows:

1. Traverse the C/A clustering tree, beginning at the trunk and sequentially examining each branching.

2. Upon arriving at a branching into subjets \( i \) and \( j \), check whether the soft drop condition is satisfied:

\[
z_{ij} > z_{cut} \left( \frac{\theta_{ij}}{R_0} \right)^\beta, \quad (4.2.3)
\]

where \( z_{cut} \) and \( \beta \) are fixed parameters of the algorithm. If so, the algorithm terminates; stop grooming and return the jet as is.

3. If the branching fails this condition, remove the softer of the two subjets (\( i \) or \( j \)) from the groomed jet and return to Step 2 on the next branching in the remaining clustering tree.

Our analysis is based on ISD where the soft drop algorithm is iterated. In this case, the procedure does not terminate when a hard branching is found, but is instead itera-
tively applied to the harder of the two subjets. This continues until an angular cutoff is reached, so in addition to \( z_{\text{cut}} \) and \( \beta \), ISD depends on an additional parameter \( \theta_{\text{cut}} \).

While ISD could be used as a grooming procedure in its own right, the primary purpose of ISD in this paper is to determine which set of \((z_{ij}, \theta_{ij})\) branchings contribute to the observables we define below. For this purpose, the ISD algorithm proceeds as follows:

1'. Set the counter \( n \) equal to 1. Traverse the C/A clustering tree, beginning at the trunk and sequentially examining each branching.

2'. Upon arriving at a branching into subjets \( i \) and \( j \), check whether the branching angle satisfies

\[
\theta_{ij} > \theta_{\text{cut}} .
\]  

(4.2.4)

If not, the algorithm terminates.

3'. If \( \theta_{ij} > \theta_{\text{cut}} \), then check whether the soft drop condition is satisfied:

\[
z_{ij} > z_{\text{cut}} \left( \frac{\theta_{ij}}{R_0} \right)^\beta .
\]  

(4.2.5)

If not, return to Step 2' on the harder of subjets \( i \) and \( j \).

4'. If the soft drop condition is satisfied, define

\[
z_n \equiv z_{ij}, \quad \theta_n \equiv \theta_{ij} .
\]  

(4.2.6)

Then increment \( n \rightarrow n + 1 \) and return to Step 2' on the harder of subjets \( i \) and
Figure 4.1: Illustration of the ISD procedure. A C/A tree is declustered from the trunk (thick line), defined by the hardest $p_T$ branches. If a node fails the soft drop condition, it is removed from consideration (dashed lines). If a node passes the soft drop condition after $n$ iterations, this defines the value of $(z_n, \theta_n)$. The declustering stops at an angular scale of $\theta_{\text{cut}}$, and subsequent nodes are not considered further (gray lines).

Because we recurse to the harder subjet at each junction, we think of each $(z_n, \theta_n)$ splitting as an emission from the “hard core” of the jet and refer to the above procedure as traversing the “trunk” of the clustering tree. A schematic of this procedure is shown in Fig. 4.1.

To emphasize, we are not using ISD as an alternative grooming technique to soft drop. In fact, we have found no need to refer to the ISD-groomed jet explicitly in our analysis. Instead, we employ ISD simply as a method to obtain an IRC-safe set of $(z_n, \theta_n)$ values to define our counting observables. Of course, the specific values of $(z_n, \theta_n)$ depend on the precise choice of ISD procedure. In this paper, we focus on the
soft drop multiplicity, which counts emissions from the trunk of the clustering tree, and have defined ISD accordingly. In Sec. 4.2.3, we consider variants of soft drop multiplicity, with corresponding variants to the ISD procedure.

To demonstrate the qualitative behavior of observables defined below in this section, we present results from parton shower simulations. We separately generate \( pp \rightarrow Z + q \) and \( pp \rightarrow Z + g \) events at center-of-mass energy 13 TeV using MadGraph 2.4.0 and let the \( Z \) decay to neutrinos for simplicity. We then shower the events through VINCIA 2.0.01 [198, 199], a plug-in to Pythia 8.215 [296], with default tuning parameters.\(^4\) Jet are identified using the anti-\( k_t \) algorithm [97] with radius \( R_0 = 0.6 \) in FastJet 3.1.3 [99]. We use a sample of events in which the hardest jet with \( |\eta| < 2.5 \) has \( p_T \) between 450 and 550 GeV. We recluster and measure our observables on the hardest jet from each event using FastJet. Because ISD is sufficiently different from ordinary soft drop, we do not use the RecursiveTools FJCONTRIB, but rather directly traverse the C/A tree in our analysis. We plan to make our code available publicly in a future release of FJCONTRIB.

\(^4\)In Sec. 4.4, we show results from four different parton shower generators. Here, we use VINCIA as a representative example since it makes predictions which are intermediate relative to the other generators.
4.2.2 Soft Drop Multiplicity

The \((z_n, \theta_n)\) values from ISD allow us to define a variety of interesting jet observables. Here, we focus on soft drop multiplicity \(n_{SD}\), which is simply the total count of the recorded \((z_n, \theta_n)\) pairs. This observable, defined already in Eq. (4.1.4), depends implicitly on the ISD parameters \(z_{\text{cut}}, \beta, \) and \(\theta_{\text{cut}}\). Among all of the observables we tested, \(n_{SD}\) appears to perform the best for quark/gluon discrimination. We discuss more general observables in Sec. 4.2.3 and App. C.1.

As defined above, ISD only follows the harder branch (i.e. the trunk) at each junction of the clustering tree. Therefore, \(n_{SD}\) effectively counts emissions from the hard core of the jet, down to the angular resolution scale \(\theta_{\text{cut}}\). When \(z_{\text{cut}} = \theta_{\text{cut}} = 0\), \(n_{SD}\) is simply the depth of the trunk of the C/A tree.

When \(z_{\text{cut}} > 0\), the soft drop multiplicity is infrared safe, as all soft emissions at finite angles fail the soft drop condition in Eq. (4.2.5). When \(\theta_{\text{cut}} > 0\), soft drop multiplicity is also collinear safe, since an exactly collinear splitting along the trunk does not satisfy Eq. (4.2.4). Alternatively, \(\beta < 0\) also gives collinear-safe distributions, since an exactly collinear splitting along the trunk does not satisfy Eq. (4.2.5). The borderline case of \(\theta_{\text{cut}} = 0\) and \(\beta = 0\) is collinear unsafe, but it can be handled using RG methods, as shown in Sec. 4.5.

In Fig. 4.2, we show the soft drop multiplicity distributions for quark and gluon
jets as extracted from VINCIA. Results are given using the benchmark parameters

\[ z_{\text{cut}} = 0.007, \quad \beta = -1, \quad \theta_{\text{cut}} = 0. \]  

(4.2.7)

This benchmark is chosen to maximize quark/gluon discrimination power while retaining perturbative calculability, as discussed in Sec. 4.3. The distributions are approximately Poisson and yield good quark/gluon discrimination power.

### 4.2.3 Multiplicity Variants

While the focus of this paper is on soft drop multiplicity \( n_{SD} \), many other observables could be defined using the \((z_n, \theta_n)\) values recorded by ISD. For example, the techniques developed in this paper can be directly applied to the weighted soft drop
\begin{equation}
n_{\text{SD}}^{(\kappa)} = \sum_{n} z_{n}^{\kappa}.
\end{equation}

Note that soft drop multiplicity is a special case ($\kappa = 0$) of this more general observable, with the same criteria for IRC safety. We study the weighted multiplicity in detail in App. C.1, but find its quark/gluon discrimination power to be inferior to the discrete $\kappa = 0$ case. In fact, LL reasoning leads one to expect the soft drop multiplicity $n_{\text{SD}}$ to have the best discrimination power of any observable defined on the $(z_{n}, \theta_{n})$ values; see the end of Sec. 4.3.3 for a short discussion.

Nevertheless, several other promising variants of soft drop multiplicity might prove useful:

- The weighted soft drop multiplicity in Eq. (4.2.8) only refers to the momentum fractions $z_{n}$ in the sum over emissions. One could also consider an angle-weighted variant

\begin{equation}
\sum_{n} z_{n}^{\kappa} \theta_{n}^{\alpha},
\end{equation}

or indeed any function of $z_{n}$ and $\theta_{n}$. The potential advantage of including $\theta_{n}$ information is that even for $\theta_{\text{cut}} = 0$, such observables would be collinear safe for $\alpha > 0$.

- Instead of counting emissions only from the trunk of the C/A tree, we could extend the sum to include all branchings down to the angular resolution $\theta_{\text{cut}}$.
This multiplicity variant would require a modification of the ISD algorithm: in step 4’, the recursion would be applied to both subjets $i$ and $j$, not just the harder one. This is a step in similarity towards full hadron multiplicity, reducing to it exactly when $z_{\text{cut}} = \theta_{\text{cut}} = 0$. This variant of soft drop multiplicity is more difficult to study analytically, however, due to the nonlinear structure of the recursion. Moreover, it is not clear that this variant would provide a performance advantage over $n_{\text{SD}}$. While gluons emitted from the hard core of a quark (gluon) jet give rise to factors of $C_F$ ($C_A$), subsequent emissions from those gluons give rise to factors of $C_A$ regardless of the jet flavor; this might wash out quark/gluon discrimination power.

- The original soft drop algorithm uses a C/A tree to mimic the angular-ordered structure of the parton shower. One could also study variants based on recluster-
ering with the generalized-$k_t$ algorithm with exponent $p$ [97, 99]. The C/A algorithm used above corresponds to $p = 0$, while the $k_t$ algorithm uses $p = 1$. For this variant, it would make sense to replace the angular cut $\theta_{\text{cut}}$ with a cut $d_{\text{cut}}$ on the generalized distance measure $d_{ij}$.

This last $k_t$ variant is of particular interest, given the discussion below in Sec. 4.3.3. Nonperturbative physics typically dominates when $k_t \simeq \Lambda_{\text{QCD}}$, so it makes sense to use a clustering algorithm where the clustering scale is “parallel” to the nonperturbative scale. This variant of $n_{\text{SD}}$ would then allow the nonperturbative phase space to
be clearly separated from the perturbative region and avoided. This would open up as much perturbative phase space for measured emissions as possible. We note that it is possible to mimic some of the LL structure of the $k_t$ variant by using ISD with $\beta = -1$, though there would be differences going to NLL order.

We defer an analysis of these variants to future work, anticipating that many of the analytic tools from this paper can be translated to these generalized contexts. Experimentally, one might want to measure a track-based version of $n_{SD}$, trading collinear safety for improved robustness to pileup, which could be studied with the help of track functions [117, 118, 165].

### 4.3 Leading-Logarithmic Analysis

At LL order, the only difference between quarks and gluons is encoded in the color factors $C_F$ and $C_A$, so Casimir scaling is a generic feature of many quark/gluon discriminants. Here, we review the case of additive observables (and close variants), where Casimir scaling of the Sudakov form factor yields a universal discrimination power at LL that depends only on $C_A/C_F$. We then show that the soft drop multiplicity is Poisson distributed, with its mean and variance satisfying Casimir scaling.

In general, any observable that is sensitive to multiple emissions at LL is “Poisson-like” distributed, in the sense that its variance $\sigma^2$ and mean $\mu$ both scale with the number $n$ of emissions counted, i.e. $\sigma^2 = O(\mu)$. In the limit of many emissions, all
such observables converge to a normal distribution with decreasing relative width\[ w_{\text{rel}} \sim \sigma/\mu \sim 1/\sqrt{n}. \] Then as more emissions are counted, the discrimination power is not a universal function of $C_A/C_F$, but instead improves as $\mu$ increases and the quark/gluon distributions separate.

In this section, we illustrate this behavior for soft drop multiplicity with distributions extracted from VINCI A, using the setup described in Sec. 4.2.1. We extract ROC (receiver operating characteristic) curves of the quark efficiency versus the gluon mistag rate, and explain their qualitative behavior. In App. C.1, we consider weighted soft drop multiplicity, with behavior that interpolates between that of Poisson- and Sudakov-distributed observables.

### 4.3.1 Review of Additive Observables

A generic jet observable is defined on the momenta $p_i$ and quantum numbers $q_i$ of particles within a jet. An additive IRC-safe observable $f$ is one that reduces to the form

\[ f\left(\{p_i, q_i\}\right) = \sum_{i \in \text{jet}} f(p_i) \tag{4.3.1} \]

in the soft/collinear limit, so that the observable depends on a simple sum over the jet constituents, independent of $q_i$.\(^5\) The function $f(p_i)$ can depend on global properties of the jet (e.g. its $p_T$), but not on its substructure. Collinear safety implies

\(^5\)One could consider additive but IRC-unsafe observables which do depend on $q_i$. 

188
that \( f(p_i) \) is linear in the particle energies \( E_i \). Examples of additive observables include the jet mass \([114, 115, 134]\), the radial moments \([189]\), and the angularities \([30, 76, 170]\), among many others.

We now review the Casimir scaling of additive observables at LL order, as discussed in Ref. \([253]\).\(^6\) For simplicity of the discussion below, we let \( \alpha_s \) be a fixed coupling so that the expressions are more compact, but it is straightforward to include a

---

\(^6\)Casimir scaling of additive observables at LL is identical to the statement of Casimir scaling of the cusp anomalous dimension in QCD, which has a long history in QCD \([64, 193]\). Casimir scaling is known to hold through three loops \([274]\) in the cusp anomalous dimension, but is not expected to hold exactly \([182]\). At NLL and beyond, Casimir scaling is broken by the appearance of the non-cusp anomalous dimension.
running coupling at LL order. At this order, we need only consider gluon emissions from the jet core that are both soft and collinear, described by the most singular terms in the splitting function. Parametrizing emissions by their angle $\theta$ and energy (or $p_T$) fraction $z$, real emissions are uniformly distributed in the $(\log 1/\theta, \log 1/z)$ plane. The density in this emission phase space is

$$\rho_i = \frac{2\alpha_s C_i}{\pi}, \quad (4.3.2)$$

where $C_i$ is the appropriate color factor, equal to $C_F = 4/3$ for quarks and $C_A = 3$ for gluons. The structure of emission phase space is shown in Fig. 4.3a. Virtual emissions are encoded in the boundaries of the emission phase space, where $\log(1/\theta), \log(1/z) \to \infty$, such that the total emission probability at each $\alpha_s$ order is zero to maintain the normalization of the probability distribution.

Applying the strongly-ordered limit and the fact that $f(p_i)$ is linear in $E_i$, only a single dominant emission contributes to the observable at lowest order:

$$\sum_{i \in \text{jet}} f(p_i) \xrightarrow{\text{LL}} \max_{i \in \text{jet}} f(p_i). \quad (4.3.3)$$

Therefore, the probability that the observable $f$ is less than some value $f_{\text{max}}$ is equal to the probability that there are no emissions in the region where $f(p_i) > f_{\text{max}}$. This implies a cumulative distribution function

$$\int_0^{f_{\text{max}}} df \, p(f) \equiv \Sigma_i(f_{\text{max}}) = e^{-\rho_i A(f_{\text{max}})}, \quad (4.3.4)$$
where \( A(f_{\text{max}}) \) is the forbidden area of emission phase space, shown in Fig. 4.3a:

\[
A(f_{\text{max}}) = \int_{f(z,\theta) > f_{\text{max}}} \frac{d\theta}{\theta} \frac{dz}{z}. \tag{4.3.5}
\]

Note that the cumulative distributions for quarks and gluons are related by

\[
\Sigma_g(f_{\text{max}}) = \left[ \Sigma_q(f_{\text{max}}) \right]^{C_A/C_F}, \tag{4.3.6}
\]

where \( C_A/C_F = 9/4 \). That is, the Sudakov form factors for \( f \) are related by Casimir scaling. As a result, the ROC curve for quark/gluon discrimination, which simply plots \( \Sigma_q(f) \) versus \( \Sigma_g(f) \), takes the universal form of Eq. (4.1.2).

From this logic, it is clear that the above analysis also extends to certain non-additive observables. For example, jet observables defined on groomed jets are not additive, since the grooming procedure removes emissions that would otherwise contribute to the sum in Eq. (4.3.1). But groomed observables of the quasi-additive form

\[
f_{\text{groomed}}(\{p_i, q_i\}) = \sum_{i \in \text{groomed jet}} f(p_i) \tag{4.3.7}
\]

still exhibit Casimir scaling, since the measured value of \( f_{\text{groomed}} \) forbid emissions in the region \( A(f_{\text{groomed}}) \) shown in Fig. 4.3b. More generally, Casimir scaling arises whenever the value of the measurement actively forbids emissions from some region of phase space. This vetoed phase space region builds up a Sudakov form factor which in turn controls the discrimination power achievable at LL.

Beyond LL order, different Sudakov-distributed observables will exhibit different discrimination power due to higher-order or nonperturbative effects, but Eq. (4.3.6)
Figure 4.4: Comparison of the quark/gluon ROC curves for various Sudakov-distributed observables to the $y = x^{9/4}$ prediction from Casimir scaling. Shown are the groomed jet radius, groomed jet mass, and ordinary jet mass. As a useful benchmark, we also show the performance of track multiplicity $n_{tr}$, which is known to be a very strong discriminant.

is still a representative benchmark. In Fig. 4.4, we show ROC curves for jet mass $m$, the soft-dropped jet mass $m_{SD}$, and the groomed jet radius $R_g$, which all roughly follow the prediction from Casimir scaling. We also show track multiplicity $n_{tr}$, which exhibits substantially better performance and provides a useful discrimination target.

### 4.3.2 Soft Drop Multiplicity

Soft drop multiplicity is not an additive observable, nor does the measured value of $n_{SD}$ actively forbid emissions in any region of phase space. As a result, $n_{SD}$ does not exhibit Sudakov behavior and it instead satisfies a fundamentally different scaling relation. Physically, this is because all emissions that pass the soft drop condition
Figure 4.5: Same as Fig. 4.3b, but now highlighting the allowed emission region $A_{\text{emit}}$ that is counted by soft drop multiplicity.

are weighted equally, so $n_{\text{SD}}$ depends on multiple emissions even at leading accuracy. These emissions occur in the region of phase space passing the soft drop and angular cuts, shown in Fig. 4.5.

Restricting to the IRC safe case with $\theta_{\text{cut}} > 0$, the measured region has finite area in the emission plane,

$$A_{\text{emit}} = \log\frac{R_0}{\theta_{\text{cut}}} \left( \log\frac{1}{2z_{\text{cut}}} + \frac{\beta}{2} \log\frac{R_0}{\theta_{\text{cut}}} \right), \quad (4.3.8)$$

and soft drop multiplicity simply counts the number of real emissions in this area.

This expression actually holds for all $\beta \in (-\infty, \infty)$ as long as the angular cut $\theta_{\text{cut}}$ imposes a non-trivial constraint on emissions. Since real emissions occur independently with uniform probability, they are described by a Poisson process, and the soft drop
multiplicity is Poisson distributed at LL order:\(^7\)

\[ P_i(n_{SD}) = \text{Pois}(\lambda_i)[n_{SD}], \quad \lambda_i = \rho_i A_{\text{emit}}. \]  \hfill (4.3.9)

For reference, the Poisson distribution with mean \( \lambda \) is

\[ \text{Pois}(\lambda)[n] = \frac{\lambda^n e^{-\lambda}}{n!}. \]  \hfill (4.3.10)

Since the variance of a Poisson distribution is also equal to \( \lambda \), the means and variances of \( n_{SD} \) both satisfy Casimir scaling

\[ \frac{\langle n_{SD} \rangle_g}{\langle n_{SD} \rangle_q} \approx \frac{C_A}{C_F}, \quad \frac{\text{Var}(n_{SD})_g}{\text{Var}(n_{SD})_q} \approx \frac{C_A}{C_F}, \]  \hfill (4.3.11)

mirroring the behavior of track multiplicity in Eq. (4.1.1), but for an IRC-safe observable.

To be clear, in defining our resummation accuracy, we count large logarithms of \( z_{\text{cut}} \) and \( \theta_{\text{cut}} \) in the mean/variance of the \( n_{SD} \) distribution. That is, we define LL and NLL exactly as for more familiar additive observables, with LL including all terms of the form \( \alpha_s^n \log^{n+1} \) that appear in the exponent of the \( n_{SD} \) distribution, and NLL including those terms of the form \( \alpha_s^n \log^n \). With this definition, Eq. (4.3.8) then shows that \( n_{SD} \) is indeed a double-logarithmic observable. In this section, we study this observable’s general properties with fixed coupling, i.e. in the double-logarithmic approximation, for purposes of illustration. In Sec. 4.4, LL and NLL results are computed using the appropriate running coupling.

\(^7\)Note that at this order, we do not account for color correlations, so the emissions are effectively Abelian.

194
The above analysis provides several concrete predictions. Our most salient result is that, since the soft drop multiplicity is Poisson distributed at LL, we expect the ratio of the variance to the mean to be close to 1, as shown in Fig. 4.6a. We also predict that the mean and variance satisfy the Casimir scaling relations in Eq. (4.3.11), as shown in Fig. 4.6b. Though not shown here, we also checked the prediction that for \( \beta = 0 \), the mean soft drop multiplicity scales as

\[
\lambda_i \propto \log \frac{1}{z_{\text{cut}}} \log \frac{1}{\theta_{\text{cut}}}.
\]

(4.3.12)

In general, we find good agreement for these predictions at large values of \( z_{\text{cut}} \), even out to \( z_{\text{cut}} \simeq 0.4 \) where \( \log z_{\text{cut}} \) is not so large. For lower cut values, nonperturbative and higher-order effects cause these LL results to break down. In Sec. 4.3.3, we demonstrate how to choose parameters so that nonperturbative effects can be avoided, and in Sec. 4.4.2, we compute the NLL corrections to the perturbative predictions discussed here.

### 4.3.3 Optimal Discrimination Power

As a direct result of the properties exhibited in Sec. 4.3.2, the discrimination power of soft drop multiplicity improves as the means \( \lambda_i = \rho_i A_{\text{emit}} \) increase. This is because the mean of each distribution is proportional to the Casimir \( C_i \), while the standard deviation is equal to the square root of the mean. The overlap of the distributions is
Figure 4.6: (a) Variance to mean ratio of the soft drop multiplicity as a function of $z_{\text{cut}}$. The parameters $\beta$ and $\theta_{\text{cut}}$ are set to the benchmark values in Eq. (4.2.7), and the LL prediction of equal mean and variance is shown as a dashed line. (b) Gluon to quark mean ratios and variance ratios, with the prediction of Casimir scaling shown as a dashed line. In both cases, we see qualitative agreement between VINCIA and the LL predictions down to $z_{\text{cut}} = 0.02$.

characterized by the relative width

$$w_{\text{rel}} \equiv \frac{\sqrt{\text{Var}(n_{\text{SD}})_i}}{\langle n_{\text{SD}} \rangle_i} = \frac{1}{\sqrt{\lambda_i}}.$$  \hspace{1cm} (4.3.13)

Indeed, in the many-emission limit where the distributions are approximately Gaussian, have equal mean and variance, and satisfy Casimir scaling, the discrimination power is solely determined by the relative width. As the cuts $z_{\text{cut}}$ and $\theta_{\text{cut}}$ are lowered, the means increase, causing the relative widths to narrow, reducing the overlap between the quark and gluon distributions, and improving the discrimination power.

For reference, the discrimination power of Poisson distributions with different means
Figure 4.7: Expected quark/gluon discrimination power for Poisson-distributed observables. The mean observable value for quarks is $\lambda_q$, and we assume the mean for gluons is given by Casimir scaling $\lambda_g = (C_A/C_F)\lambda_q$. For reference, we show the $y = x^{9/4}$ curve for additive observables with Casimir scaling, as well as track multiplicity $n_t$ extracted from VINCIA. For mean quark values $\lambda_q \gtrsim 2$, a Poisson-like observable satisfying Casimir scaling would be competitive with track multiplicity. The ROC curves are piecewise linear since the observable takes on discrete integer values.

is shown in Fig. 4.7, from which we see that track multiplicity has comparable discrimination power to a $\lambda_q \simeq 2$ observable.

To maximize the quark/gluon discrimination power, one should maximize the mean of the soft drop multiplicity distributions, which corresponds to taking $z_{\text{cut}}$ and $\theta_{\text{cut}}$ as small as possible, for a given exponent $\beta$. The validity of this analysis, however, is restricted to perturbation theory, so we must ensure that the values of the chosen parameters do not allow for distributions that are dominated by nonperturbative emissions. We can determine the parameters that enforce perturbative emissions by
Figure 4.8: Illustration of the optimal phase space configuration consistent with a perturbative analysis. The dashed line with slope $-1$ separates perturbative and nonperturbative emissions. (a) For $\beta > -1$, the value of $\theta_{\text{cut}}$ has to be chosen to avoid allowed emissions above the nonperturbative boundary. (b) For $\beta < -1$, $\theta_{\text{cut}}$ can be set to zero, with $z_{\text{cut}}$ pushed to the nonperturbative boundary. To maximize the allowed perturbative phase space, one should take $\beta = -1$ and $z_{\text{cut}}$ set to the optimal value in Eq. (4.3.19).

restricting the minimum relative $k_t$ appropriately.

To enforce that an emission is perturbative, we require that the relative $k_t$ of the emission is larger than a perturbative cutoff scale $\Lambda_{\text{NP}}$, i.e.

$$z \theta \gtrsim \frac{\Lambda_{\text{NP}}}{p_T},$$

where $z$ and $\theta$ are the energy fraction and splitting angle of the emission, and $p_T$ is the transverse momentum of the jet. Below, we take $\Lambda_{\text{NP}} = 2$ GeV unless otherwise noted. For an emission that just passes soft drop, and therefore contributes to the soft
drop multiplicity, we have
\[ z \gtrsim z_{\text{cut}} \frac{\theta^\beta}{R_0^\beta}. \]  

(4.3.15)

There are two regimes to consider. For $\beta > -1$ as in Fig. 4.8a, we can find the intersection of Eqs. (4.3.14) and (4.3.15). Setting $\theta \to \theta_{\text{cut}}$, we find a restriction on $\theta_{\text{cut}}$ to be perturbative:
\[ \theta_{\text{cut}} \gtrsim \left( \frac{\Lambda_{\text{NP}}}{z_{\text{cut}} p_T R_0} \right)^{\frac{1}{1+\beta}} R_0. \]  

(4.3.16)

To determine the optimal choice of $z_{\text{cut}}$ while enforcing perturbativity, we set $\theta_{\text{cut}}$ to saturate this inequality and insert it into the double-log expression for the average soft drop multiplicity, Eq. (4.3.8). Maximizing this quantity, we find the optimal ISD parameters to be
\[ z_{\text{cut,\;optimal}} = \frac{1}{2} \left( \frac{2 \Lambda_{\text{NP}}}{p_T R_0} \right)^{\frac{1}{1+\beta}}, \]  

(4.3.17)
\[ \theta_{\text{cut,\;optimal}} = \left( \frac{2 \Lambda_{\text{NP}}}{p_T R_0} \right)^{\frac{1}{1+\beta}} R_0. \]  

(4.3.18)

The factors of two arise because the energy fraction of the softer emission is (by definition) less than 1/2. Inserting these results into the expression for the average soft drop multiplicity, we find the largest perturbative value for the mean soft drop multiplicity to be
\[ \langle n_{\text{SD}} \rangle_{\text{optimal}}^{\beta>1} \simeq \frac{\alpha_s}{\pi} \frac{C_t}{2 + \beta} \log^2 \left( \frac{2 \Lambda_{\text{NP}}}{p_T R_0} \right). \]  

(4.3.19)

For $\beta < -1$, one can see from the $(\log 1/\theta, \log 1/z)$ phase space in Fig. 4.8b that
an angular cutoff is not needed to avoid the nonperturbative region, so we can set \( \theta_{\text{cut}} = 0 \). In this case, \( z_{\text{cut}} \) saturates the bound Eq. (4.3.14) for \( \theta \rightarrow R_0 \), yielding

\[
z_{\text{cut}}|_{\text{optimal}} = \frac{\Lambda_{\text{NP}}}{p_T R_0},
\]  

(4.3.20)

and the average soft drop multiplicity is

\[
\langle n_{SD}\rangle_{\beta<-1}^{\text{optimal}} \simeq \frac{\alpha_s C_i}{\pi |\beta|} \log^2 \left( \frac{2\Lambda_{\text{NP}}}{p_T R_0} \right).
\]  

(4.3.21)

Combining these regions for all \( \beta \in (-\infty, \infty) \), the maximum attainable mean soft drop multiplicity with perturbative parameters is

\[
\langle n_{SD}\rangle_{\text{optimal}} \simeq \frac{\alpha_s C_i}{\pi} \min \left[ \frac{1}{|\beta|}, \frac{1}{2 + |\beta|} \right] \log^2 \left( \frac{2\Lambda_{\text{NP}}}{p_T R_0} \right).
\]  

(4.3.22)

In particular, the mean is maximized for \( \beta = -1 \), giving the optimal perturbative discrimination power in this double-log approximation. This result can be understood directly from Fig. 4.8, which shows that soft drop multiplicity with \( \beta = -1 \) can capture all of the perturbative emissions in phase space.

We can directly test this double-log prediction in parton shower generators. In Fig. 4.9a, we show the quark/gluon ROC curve for soft drop multiplicity with the optimal perturbative soft drop parameters, sweeping through \( \beta \). The best discrimination power found in VINCI is indeed observed near \( \beta = -1 \). For a more quantitative test, Eq. (4.3.22) predicts that the ratio of the optimal soft drop multiplicity for a given value of \( \beta \) to the optimal soft drop multiplicity at \( \beta = 0 \) is

\[
\frac{\langle n_{SD}\rangle_{\text{optimal}}}{\langle n_{SD}\rangle_{\beta=0}^{\text{optimal}}} = \min \left[ \frac{2}{|\beta|}, \frac{2}{2 + |\beta|} \right].
\]  

(4.3.23)

200
Figure 4.9: (a) Discrimination power of soft drop multiplicity as a function of \( \beta \), with the optimal (perturbative) values of \( z_{\text{cut}} \) and \( \theta_{\text{cut}} \) computed from Eqs. (4.3.17), (4.3.18), and (4.3.20) using \( \Lambda_{\text{NP}} = 2 \) GeV. (b) Ratio of mean \( n_{\text{SD}} \) as a function of \( \beta \) to mean \( n_{\text{SD}} \) at \( \beta = 0 \). The VINCIA results for quarks and gluons agree with the double log prediction from Eq. (4.3.22), except near \( \beta = -1 \) where nonperturbative effects become important.

In Fig. 4.9b, we compare this ratio to distributions extracted from VINCIA and find good agreement away from \( \beta = -1 \). Note that when \( \beta = -1 \), the counted and nonperturbative regions share a boundary, while in all other cases the two regions only meet at a single point. This explains why nonperturbative sensitivity should be amplified when \( \beta \) nears \( -1 \). This extra sensitivity could of course be mitigated by using a more conservative value of \( \Lambda_{\text{NP}} \), but there is a tradeoff between reducing nonperturbative effects and increasing discrimination power.

In Fig. 4.10a, we show the effect that decreasing \( \Lambda_{\text{NP}} \) (and thus decreasing \( z_{\text{cut}} \) and
Figure 4.10: (a) Discrimination power of soft drop multiplicity as a function of $\Lambda_{\text{NP}}$ with $\beta = -1$, $\theta_{\text{cut}} = 0$, and $z_{\text{cut}}$ computed from Eq. (4.3.20). (b) Impact of hadronization and underlying event in VINCIA on gluon distributions.

$\theta_{\text{cut}}$) has on the discrimination power, holding $\beta = -1$ fixed. Note that $n_{\text{SD}}$ rivals $n_{\text{tr}}$ for $\Lambda_{\text{NP}} = 1 \text{ GeV}$, but that there is no gain in performance when $\Lambda_{\text{NP}}$ is taken smaller. In Fig. 4.10b we show the shift in gluon $n_{\text{SD}}$ distributions from switched off hadronization and underlying event in VINCIA. We take this as an indicator of non-perturbative sensitivity in the distributions. One can see that perturbative control is lost for $\Lambda_{\text{NP}} < 2 \text{ GeV}$. For $p_T = 500 \text{ GeV}$, $\Lambda_{\text{NP}} = 2 \text{ GeV}$ gives the benchmark parameters in Eq. (4.2.7).

Our perturbative analysis here was restricted to LL order and fixed coupling, and the inclusion of higher-order effects will affect the discrimination power of soft drop multiplicity. In particular, at NLL order, quark and gluon jet flavors can mix, so we
expect that higher-order effects in general decrease the discrimination power from the LL prediction. We perform NLL calculations and compare our results to parton showers in Sec. 4.4. Beyond these higher-order effects, we have restricted the analysis to perturbative parameters. Allowing nonperturbative emissions to contribute to the soft drop multiplicity should improve the discrimination power, however, at the expense of loss of predictivity. We discuss in Sec. 4.5 how to restore some of this predictive power in the nonperturbative regime with GFFs.

One might wonder if the discrimination power could be further improved by weighting the emissions, e.g. by their energy, as in the weighted soft drop multiplicity of Eq. (4.2.8). At LL order, however, the soft drop multiplicity is provably the most powerful discriminant that can be defined on the \((z_n, \theta_n)\) values. To see this, note that the normalized distribution of emissions in the \((\log 1/\theta, \log 1/z)\) plane is identical for quark and gluon jets at LL order, even including running coupling effects. Therefore, once the value of \(n_{SD}\) is known for a given jet, no additional discriminatory information can be gleaned from the \((z_n, \theta_n)\) values. Nevertheless, weighted soft drop multiplicity provides an example of a more general observable that can be effectively studied with our analytic tools; we demonstrate this in App. C.1.9

---

8We thank Ben Nachman for discussions on this point. Specifically, he demonstrated that the quark/gluon likelihood ratio is a monotonic function of \(n_{SD}\), with no other non-trivial \((z_n, \theta_n)\) dependence, thus providing further confirmation that \(n_{SD}\) is the optimal discriminant one can construct.

9One might be attracted to weighted soft drop multiplicity because it reduces sensitivity to soft emissions. Presumably, the value of \(\Lambda_{NP}\) could be reduced somewhat without introducing
4.4 Calculations for IRC-Safe Soft Drop Multiplicity

We now demonstrate that the LL predictions of the previous section can be reproduced by a set of perturbative evolution equations. These equations describes how soft drop multiplicity evolves with decreasing \( \theta_{\text{cut}} \), similar to traditional parton evolution [31]. This approach also admits a generalization to NLL, which we use to make precise predictions for comparison to parton showers.

When talking about the resummation of large logarithms at LL and NLL accuracy, we are specifically referring to factors of \( \log z_{\text{cut}} \) and \( \log \theta_{\text{cut}} \), not to any logarithms associated with the \( n_{\text{SD}} \) observable (which is an integer). As we already saw in Eq. (4.3.8), these logarithms control the size of the emission phase space, which in turn control the expected mean value of \( n_{\text{SD}} \), so their resummation is essential for predicting the distribution of \( n_{\text{SD}} \).

4.4.1 Leading-Logarithmic Evolution Equations

We begin by analyzing the soft drop multiplicity to LL accuracy. This case is simple enough to keep the structure of the \( \theta_{\text{cut}} \) evolution transparent; the generalization

significant nonperturbative effects. One cannot increase perturbative discrimination power in this way, however, since any gain in discrimination power from reducing \( \Lambda_{\text{NP}} \) must necessarily come with comparable nonperturbative sensitivity.
to NLL just requires keeping track of more details. To achieve LL accuracy, we need only consider soft-collinear gluons emitted from the hard core of a jet; flavor-changing effects are not present at this order. Furthermore, the trunk of the clustering tree retains all but an $\mathcal{O}(z_{\text{cut}})$ fraction of the original jet’s energy, so for $z_{\text{cut}} \ll 1$, energy losses are negligible at this order as well.

Let $p_n^i(\theta_{\text{cut}})$ denote the probability that, given a jet of flavor $i$ and ISD parameter $\theta_{\text{cut}}$, its soft drop multiplicity $n_{\text{SD}}(\theta_{\text{cut}})$ is measured to be $n$. Here, we leave the dependence on $z_{\text{cut}}$ and $\beta$ implicit, since they do not participate directly in the evolution equations. Since $n_{\text{SD}}$ is a discrete counting observable, $p_n^i(\theta_{\text{cut}})$ is finite and should satisfy the normalization condition $\sum_{n=0}^{\infty} p_n^i(\theta_{\text{cut}}) = 1$ for each flavor $i$.

We can compute the distribution for $p_n^i(\theta_{\text{cut}})$ by solving a set of evolution equations. Consider decreasing the resolution angle from $\theta_{\text{cut}}$ to $\theta_{\text{cut}} - \delta \theta_{\text{cut}}$. The value of $n_{\text{SD}}$ will increase by one if there is an emission in the interval $[\theta_{\text{cut}} - \delta \theta_{\text{cut}}, \theta_{\text{cut}}]$ that passes soft drop; otherwise $n_{\text{SD}}$ will remain unchanged. That is,

$$p_n^i(\theta_{\text{cut}} - \delta \theta_{\text{cut}}) = p_{n-1}^i(\theta_{\text{cut}}) \frac{\delta \theta_{\text{cut}}}{\theta_{\text{cut}}} \int_0^{1/2} dz \frac{\alpha_s(z \theta_{\text{cut}} p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{\text{SD}}(z, \theta_{\text{cut}})$$

$$+ p_n^i(\theta_{\text{cut}}) \left( 1 - \frac{\delta \theta_{\text{cut}}}{\theta_{\text{cut}}} \int_0^{1/2} dz \frac{\alpha_s(z \theta_{\text{cut}} p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{\text{SD}}(z, \theta_{\text{cut}}) \right).$$

(4.4.1)

Here, $P_{i \rightarrow i}(z)$ is the splitting function for the hard parton $i$ to emit a collinear gluon of energy fraction $z$ (and remain as flavor $i$), and $\Theta_{\text{SD}}(z, \theta)$ imposes the soft drop con-
condition,
\[ \Theta_{SD}(z, \theta) \equiv \Theta \left( z - z_{cut} \frac{\theta^\beta}{R_0^\beta} \right). \]  \hspace{1cm} (4.4.2)

At LL, \( \alpha_s(z \theta cut p_T) \) runs with the 1-loop \( \beta \) function.

Using Eq. (4.4.1), we can derive the linear first-order differential equation in \( \theta_{cut} \),
\[ \frac{d p_n^i(\theta_{cut})}{d \theta_{cut}} = p_n^i(\theta_{cut}) - p_{n-1}^i(\theta_{cut}) \int_0^{1/2} dz \frac{\alpha_s(z \theta_{cut} p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{SD}(z, \theta_{cut}). \]  \hspace{1cm} (4.4.3)

Because no emissions are recorded outside the jet radius \( R_0 \), there is a boundary condition \( p_n^i(R_0) = \delta_{n,0} \). With this boundary condition, the solution to Eq. (4.4.3) is
\[ p_0^i(\theta_{cut}) = e^{-I_{i \rightarrow i}(\theta_{cut}, R_0)}, \]  \hspace{1cm} (4.4.4)
\[ p_{n \geq 1}^i(\theta_{cut}) = \int_{\theta_{cut}}^{R_0} \frac{d \theta}{\theta} e^{-I_{i \rightarrow i}(\theta_{cut}, \theta)} \left( \int_0^{1/2} dz \frac{\alpha_s(z \theta p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{SD}(z, \theta) \right) p_{n-1}^i(\theta), \]  \hspace{1cm} (4.4.5)

where
\[ I_{i \rightarrow i}(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} \frac{d \theta}{\theta} \int_0^{1/2} dz \frac{\alpha_s(z \theta p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{SD}(z, \theta). \]  \hspace{1cm} (4.4.6)

The expression in Eq. (4.4.4) corresponds to the case of no emissions between \( R_0 \) and \( \theta_{cut} \). The expression in Eq. (4.4.5) computes the probability that ISD records \( n - 1 \) emissions in the interval \( [\theta, R_0] \), one final emission at \( \theta \), then zero emissions in the interval \( [\theta_{cut}, \theta] \), with an integral over the angle \( \theta \) where the final counted emission occurs.

We can interpret Eq. (4.4.5) as a recursion relation in \( n \) with Eq. (4.4.4) as the ini-
tial condition. The first step in the recursion \((n = 1)\) gives
\[
 p_1^i(\theta_{\text{cut}}) = \int_{\theta_{\text{cut}}}^{R_0} \frac{d\theta}{\theta} e^{-I_{i \rightarrow i}(\theta_{\text{cut}}, \theta)} \left( \int_0^{1/2} dz \frac{\alpha_s(z \theta \rho_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{\text{SD}}(z, \theta) \right) e^{-I_{i \rightarrow i}(\theta, R_0)}
\]
\[
 = e^{-I_{i \rightarrow i}(\theta_{\text{cut}}, R_0)} I_{i \rightarrow i}(\theta_{\text{cut}}, R_0). \tag{4.4.7}
\]

A similar simplification occurs for each value of \(n\), and we recognize the Poisson distribution we found in Eq. (4.3.9):
\[
 p_n^i(\theta_{\text{cut}}) = \frac{1}{n!} \left[ I_{i \rightarrow i}(\theta_{\text{cut}}, R_0) \right]^n e^{-I_{i \rightarrow i}(\theta_{\text{cut}}, R_0)}. \tag{4.4.8}
\]

At LL, the soft drop multiplicity \(n_{\text{SD}}\) is thus Poisson distributed with mean \(I_{i \rightarrow i}(\theta_{\text{cut}}, R_0)\). With fixed coupling, the mean value agrees exactly with \(\lambda_i = \rho_i A_{\text{emit}}\) found before (see Eqs. (4.3.2) and (4.3.8)):
\[
 I_{i \rightarrow i}(\theta_{\text{cut}}, R_0) |_{\text{fixed } \alpha_s} = \frac{2\alpha_s C_i}{\pi} \log \frac{R_0}{\theta_{\text{cut}}} \left( \log \frac{1}{2z_{\text{cut}}} + \frac{\beta}{2} \log \frac{R_0}{\theta_{\text{cut}}} \right). \tag{4.4.9}
\]

### 4.4.2 Next-to-Leading-Logarithmic Corrections

The next-to-leading logarithms take the form \(\alpha_s^n \log^n z_{\text{cut}}\) and \(\alpha_s^n \log^n \theta_{\text{cut}}\) in the logarithm of \(p_n^i(\theta_{\text{cut}})\). To resum these, we must consider emitted partons that are not necessarily soft and that can be either quarks or gluons. This requires us to take energy losses and flavor changes into account at this accuracy. It is convenient to compute \(p_n^i(\theta_{\text{cut}})\) by expressing it as
\[
p_n^i(\theta_{\text{cut}}) = \sum_{j=q,g} \int_{1/2^n}^1 dZ p_{n}^{i \rightarrow j}(Z)(\theta_{\text{cut}}). \tag{4.4.10}
\]
Here, $dZ p_{n}^{i \rightarrow j(Z)}(\theta_{\text{cut}})$ is the differential probability that, given a jet of flavor $i$, ISD counts $n$ emissions from its hard core that result in a flavor change from $i$ to $j$, and a remaining energy fraction in the interval $[Z, Z + dZ]$. These more differential distributions evolve with $\theta_{\text{cut}}$ as

$$
\begin{align*}
  p_{n}^{i \rightarrow j(Z)}(\theta_{\text{cut}} - \delta\theta_{\text{cut}}) &= p_{n}^{i \rightarrow j(Z)}(\theta_{\text{cut}}) \left( 1 - \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} \int_{0}^{1/2} dz \frac{\alpha_s(z \theta_{\text{cut}} Z_p r)}{\pi} P_{j \rightarrow \text{any}}(z) \Theta_{\text{SD}}(z, \theta_{\text{cut}}) \right) \\
  + \sum_{k} \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} \int_{0}^{1/2} dz \frac{\alpha_s(z \theta \frac{Z}{1-z} p_r)}{\pi} P_{k \rightarrow j}(z) \Theta_{\text{SD}}(z, \theta) p_{n-1}^{i \rightarrow k[(1-z)]}(\theta_{\text{cut}}) \frac{1}{1 - z},
\end{align*}
$$

(4.4.11)

where

$$
P_{i \rightarrow \text{any}}(z) = \sum_{j} P_{i \rightarrow j}(z).
$$

(4.4.12)

The middle line of Eq. (4.4.11) is the probability that $n$ emissions are counted at resolution $\theta_{\text{cut}}$, and that only virtual or soft-dropped emissions (neither of which have an impact on energy fractions, up to $z_{\text{cut}}$ corrections) occur in the interval $[\theta_{\text{cut}} - \delta\theta_{\text{cut}}, \theta_{\text{cut}}]$. The second line is the probability that $n - 1$ emissions are counted at resolution $\theta_{\text{cut}}$ and result in a flavor conversion $i \rightarrow k$, and that an additional counted emission causing further conversion $k \rightarrow j$ occurs in $[\theta_{\text{cut}} - \delta\theta_{\text{cut}}, \theta_{\text{cut}}]$.

We now justify that these evolution equations do indeed resum large logarithms to NLL, with one caveat. As is necessary for NLL resummation, these evolution equa-

---

10 The probability for the hard core to be left with energy fraction between $Z/(1 - z)$ and $(Z + dZ)/(1 - z)$ is then $p_{n}^{i \rightarrow j[(1-z)]}(\theta_{\text{cut}}) dZ/(1 - z)$. This is used in Eq. (4.4.11).
tions contain NLO information about the jet’s substructure. To achieve NLL accuracy, we need to properly include the following double-emissions structures: collinear plus collinear (C+C), soft plus collinear (S+C), soft plus soft (S+S), and hard plus soft-collinear (H+SC). Since ISD is an angular-ordered algorithm, collinear emissions factorize in the cross section, so our evolution equations correctly include C+C and S+C double emissions. The S+S case is included as well by letting $\alpha_s$ run with the 2-loop $\beta$ function in the CMW scheme [116]. The one caveat is that we do not describe H+SC double emissions correctly at NLO, since we use splitting functions instead of full matrix elements.\textsuperscript{11} Thus, our approximation should become more accurate as the jet radius $R_0$ becomes smaller, forcing hard emissions in the jet to become collinear. We also ignore the effects of logarithms of $z_{cut}$ that arise from nonglobal radiation [151], and so do not describe emissions in the jet from secondary radiation from outside of the jet.

Despite the extra complications at NLL order, Eq. (4.4.11) is still a linear first-order differential equation, just as in Sec. 4.4.1. The solution is

$$
p_{0 \to j(Z)}(\theta_{cut}) = \delta_{ji} \delta(Z - 1) \exp[-I_{1 \to \text{any}}(\theta_{cut}, R_0)],
$$

$$
p_{n \geq 1 \to j(Z)}(\theta_{cut}) = \sum_k \int_{\theta_{cut}}^{R_0} \frac{d\theta}{\theta} \int_0^{1/2} dz \exp[-I_{j(Z) \to \text{any}}(\theta_{cut}, \theta)]
\times \frac{\alpha_s(z \theta \frac{Z}{1-z} p_T)}{\pi} P_{k \to j}(z) \Theta_{SD}(z, \theta) P_{n-1 \to k[Z/(1-z)]}(\theta) \frac{1}{1-z},
$$

\text{\textsuperscript{11}Besides this caveat, though, note that our use of } 1 \rightarrow 2 \text{ (as opposed to } 1 \rightarrow 3 \text{) splitting functions is sufficient at NLL, since } n_{SD} \text{ is a double-logarithmic observable.}
where

\[ I_{j(Z)\rightarrow \text{any}}(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} d\theta \int_0^{1/2} dz \frac{\alpha_s(z \theta Z p_T)}{\pi} P_{j\rightarrow \text{any}}(z) \Theta_{SD}(z, \theta). \]  
(4.4.15)

Note that \( p_{n}^{i\rightarrow j(Z)} \) vanishes for \( Z < 1/2^n \). The same manipulations that led to Eq. (4.4.7) and the Poisson distribution at LL do not go through at NLL, so we cannot write \( p_{n}^{i\rightarrow j(Z)} \) or \( p_{n}^{i} \) in closed form at this order. Nonetheless, the integrals in Eq. (4.4.14) can be performed numerically by first computing \( p_{1}^{i\rightarrow j(Z)} \), then computing \( p_{2}^{i\rightarrow j(Z)} \), and so on until \( p_{n}^{i} \) is negligible. In practice, the probability saturates for \( n \) of order 10.

The \( n_{SD} \) distributions and ROC curves at LL and NLL accuracy are displayed in Fig. 4.11. The uncertainties in the NLL calculation come from varying the \( \alpha_s \) scale up and down by a factor of 2. (Scale variation in the LL calculation does not give a reliable estimate of the uncertainty, since flavor-changing processes are absent at LL; we therefore omit bands around the LL predictions.) The fact that the uncertainties are abnormally small in one bin is an artifact of this one-dimensional variation procedure, which leaves the scale-varied distributions properly normalized. Also, the uncertainties in the ROC curve are substantially smaller than the uncertainties in the NLL distributions, since the way we implement the scale variation affects quarks and gluons in a correlated way. We show both \( \beta = -1 \) and \( \beta = -0.5 \) with \( z_{\text{cut}} \) and \( \theta_{\text{cut}} \) chosen to be “optimal” according to Eqs. (4.3.17), (4.3.18), and (4.3.20) with \( \Lambda_{NP} = 2 \text{ GeV} \). One can see that NLL corrections result in a slight decrease in discrimination power.
Figure 4.11: Calculations at LL and NLL accuracy for (left column) $n_{SD}$ distributions and (right column) the corresponding quark/gluon ROC curves. Parameters are chosen according to Eqs. (4.3.17), (4.3.18), and (4.3.20) with $\Lambda_{NP} = 2$ GeV and (top row) $\beta = -1$ and (bottom row) $\beta = -0.5$. The uncertainties in the NLL calculation come from varying the $\alpha_s$ scale by a factor of 2.
compared to LL, due in part to the flavor changes that occur at this order.

4.4.3 Comparison to Parton Showers

It is instructive to compare our NLL calculation of the soft drop multiplicity $n_{SD}$ with results obtained from parton shower generators. In addition to the VINCIA setup described in Sec. 4.2.1, we obtained alternative event samples by showering the hard events through PYTHIA 8.219 [2, 295], HERWIG 7.0.1 [41, 73], and SHERPA 2.2.0 [200], interfaced to their default hadronization and underlying event models.

First, to validate the reliability of our NLL calculation, we want to explore the impact of nonperturbative effects on the parton showers. In Sec. 4.3.3 we noted that hadronization effects should generically be minimal provided parameters are chosen at or above the values given in Eqs. (4.3.17), (4.3.18), and (4.3.20). To investigate this expectation further, we check the size of nonperturbative corrections in VINCIA by turning hadronization and underlying event off and comparing to results obtained using the default settings. In Fig. 4.12, we show $n_{SD}$ with $\beta = -1$ and $\beta = -0.5$, where in each case $z_{cut}$ and $\theta_{cut}$ are computed using Eqs. (4.3.17), (4.3.18), and (4.3.20) with $\Lambda_{NP} = 2$ GeV. As expected, nonperturbative effects are under control, confirming that our perturbative NLL calculations should indeed be reliable in predicting the $n_{SD}$ distributions. Though not shown, the other three parton shower generators also exhibit comparable nonperturbative shifts.
Figure 4.12: Impact of nonperturbative effects on (left column) $n_{SD}$ distributions and (right column) the corresponding ROC curves. This study employs VINCIA, where parameters are chosen according to Eqs. (4.3.17), (4.3.18), and (4.3.20) with $\Lambda_{NP} = 2$ GeV and (top row) $\beta = -1$ and (bottom row) $\beta = -0.5$. 

213
Figure 4.13: Predicted quark/gluon discrimination power from (a) **Pythia 8.219**, (b) **Herwig 7.0.1**, (c) **Sherpa 2.2.0**, and (d) **Vincia 2.0.01**. While the generators disagree about absolute performance, they agree that $n_{SD}$ with $\beta = -1$ outperforms jet mass and approaches the discrimination power of $n_{tr}$.
Figure 4.14: Analytic NLL distributions compared to parton shower generators for (top row) quark jets, (middle row) gluon jets, along with (bottom row) the corresponding ROC curves. Parameters are chosen according to Eqs. (4.3.17), (4.3.18), and (4.3.20) with $\Lambda_{NP} = 2$ GeV and (left column) $\beta = -1$ and (right column) $\beta = -0.5$. 
Next, we show that all parton shower generators predict that soft drop multiplicity is a relatively good quark/gluon discriminant. In Fig. 4.13, we compare \( n_{\text{SD}} \) with \( \beta = -1 \) and \( \beta = -0.5 \) to jet mass and \( n_{\text{tr}} \) for each generator separately. For \( \beta = -1 \), soft drop multiplicity provides a significant improvement over generic additive observables but does not quite match the performance of track multiplicity. (See, however, Fig. 4.10a where nonperturbative parameter values push the performance of \( n_{\text{SD}} \) to match \( n_{\text{tr}} \).) The ordering of the ROC curves is roughly the same between the four generators, though the absolute discrimination power does differ.

Finally, we can directly compare our NLL predictions to the parton shower generators. In Fig. 4.14, we show the \( n_{\text{SD}} \) distributions and ROC curves for both \( \beta = -1 \) and \( \beta = -0.5 \). When interpreting these curves, one has to remember that the NLL prediction does not include nonperturbative effects. The quark distributions are roughly similar between the various generators, but there is a larger spread in the gluon distributions, a feature also seen in the study of Refs. [36, 205]. It is interesting to note that both VINCIA and SHERPA, as well as our NLL calculation, predict rather strong discrimination power, in better agreement with PYTHIA than with HERWIG. This highlights the importance of carrying out these analytic calculations to even higher accuracy, in order to better understand the desired behavior for these parton shower generators.
4.5 Calculations for Collinear-Unsafe Soft Drop

Multiplicity

Thus far, we have focused on choices of ISD parameters where the quark/gluon discrimination power could be predicted using perturbation theory. In the section, we consider the special case of $\theta_{\text{cut}} = 0$ and $\beta = 0$, where the soft drop multiplicity is collinear unsafe but still soft safe, allowing us to calculate its RG evolution.

4.5.1 Review of Generalized Fragmentation Functions

To study observables with purely collinear final-state divergences, one can use the formalism of GFFs. Ordinary fragmentation functions are well-known objects in QCD which describe the fragmentation of a quark or gluon into a single hadron. GFFs are nonperturbative objects that describe the fragmentation of a quark or gluon into correlated sets of hadrons. The GFF technique has already been applied successfully to weighted jet charge [243, 310], track functions [117, 118], and generalized angularities [255], and a forthcoming paper explores the broader space of observables described by GFFs [165].

Each collinear-unsafe observable $x$ has an associated set of GFFs, $F_i(x, \mu)$, where $i$ labels each quark flavor, anti-quark flavor, and gluon. They are normalized to have
unit integral,
\[
\int_{-\infty}^{\infty} dx \mathcal{F}_i(x, \mu) = 1,
\]
and at leading order, they have the interpretation of the probability of parton \( i \) to yield the observable value \( x \). In higher-order partonic calculations, the GFFs absorb collinear divergences and pick up dependence on the RG scale \( \mu \). While the GFFs themselves cannot be calculated using perturbation theory, their RG evolution is calculable. Ordinary fragmentation functions exhibit linear DGLAP evolution [31, 161, 210], whereas GFFs in general have non-linear evolution equations which can even involve mixing between different sets of GFFs.

As shown in Ref. [165], though, for observables defined on a pairwise clustering tree, the evolution equations for the GFFs greatly simplify. These observables are called fractal jet observables, since their RG evolution is reminiscent of the fractal structure of the parton shower. For \( \theta_{\text{cut}} = 0 \) and \( \beta = 0 \), soft drop multiplicity (and its weighted variant) is an example of a fractal jet observable, allowing us to use the GFF formalism.

It is important to emphasize that the GFF formalism only works for purely collinear divergences. For \( \theta_{\text{cut}} = 0 \) but \( \beta > 0 \), there are mixed soft-collinear divergences in the simultaneous \( z \to 0 \) and \( \theta \to 0 \) limits. These correlated diverges would require additional regulators, similar in spirit to rapidity regularization [131] (see also [83]). The use of fragmentation functions to study the \( \beta = 0 \) limit was previously considered in
Ref. [249] to study the soft-dropped $z_g$ distribution (which is the same as $z_1$ for ISD).

Following Ref. [165], consider a fractal observable $x$ defined recursively on an IRC-safe binary clustering tree as follows. Each final-state hadron is assigned a starting weight $w_n$, which serves as the initial seed for the observable, and the observable $x$ is built recursively according to

$$x = \hat{x}(z, x_1, x_2), \quad (4.5.2)$$

where $z \in [0, 1]$ is the momentum fraction of the $2 \to 1$ merging, and $x_1$ and $x_2$ are the values of the observable (or the starting weight $w_n$) on the daughter nodes. Note that $\hat{x}$ is independent of the opening angle $\theta$ of the merging, and the only angular dependence comes through the choice of clustering tree. The leading-order RG evolution for the GFFs associated with $x$ is

$$\mu \frac{d}{d\mu} \mathcal{F}_i(x, \mu) = \frac{1}{2} \sum_{jk} \int dz \, dx_1 \, dx_2 \frac{\alpha_s(\mu)}{\pi} P_{i \to jk}(z) \mathcal{F}_j(x_1, \mu) \mathcal{F}_k(x_2, \mu) \delta [x - \hat{x}(z, x_1, x_2)], \quad (4.5.3)$$

where $P_{i \to jk}(z)$ is the splitting function. At this order, the evolution equation (but not the observable itself) is independent of the choice of clustering tree. Note that the evolution equation is also independent of the starting weights $w_n$, which are effectively encoded in the low-scale initial conditions for $\mathcal{F}_i$. Even though the clustering tree is IRC safe, $x$ is generally collinear unsafe, since Eq. (4.5.2) allows an exactly collinear splitting to change the observable.
The canonical RG scale for a generic GFF is

$$\mu = E_{\text{jet}} R_0,$$

(4.5.4)

and if we can extract the functional form of $F_s(x, \mu)$ at a low scale, we can use Eq. (4.5.3) to predict their form at a higher scale. The RG equations have the same recursive structure as a parton shower, and we can use the numerical techniques of Ref. [165] to evolve the GFFs in $\mu$. As we will see, our observable of interest actually has a linear evolution equation, which greatly simplifies the numerical treatment.

### 4.5.2 Linear Evolution for Soft Drop Multiplicity

For $\theta_{\text{cut}} = 0$ and $\beta = 0$, soft drop multiplicity is an example of a fractal observable. More generally, any ISD observable of the form

$$x = \sum_{n} f(z_n)$$

(4.5.5)

is a fractal observable. Using C/A for the binary clustering tree with starting weights $w_a = 0$, the recursion relation for this general observable is

$$\hat{x}(z, x_1, x_2) = \begin{cases} 
    x_2 & 0 \leq z < z_{\text{cut}}, \\
    x_2 + f(z) & z_{\text{cut}} \leq z \leq 1/2, \\
    x_1 + f(1-z) & 1/2 \leq z \leq 1 - z_{\text{cut}}, \\
    x_1 & 1 - z_{\text{cut}} < z \leq 1.
\end{cases}$$

(4.5.6)
The four cases check which subjet is harder and whether the softer subjet passes soft drop. If the softer subjet fails soft drop (i.e. \( \min(z, 1 - z) < z_{\text{cut}} \)), then the observable value is unchanged. If the softer subjet passes soft drop, then the \( f(z) \) (or \( f(1 - z) \)) value of the splitting enters linearly into the observable.

The recursion relation in Eq. (4.5.6) takes a particularly simple form, since each of the four cases involves either \( x_1 \) or \( x_2 \), but not both. This allows us to rewrite the RG evolution from Eq. (4.5.3) in the form

\[
\mu \frac{dF_i(x, \mu)}{d\mu} = \sum_{jk} \frac{\alpha_s(\mu)}{\pi} \left( \int_{0}^{z_{\text{cut}}} dz P_{i \to jk}(z) F_k(x, \mu) + \int_{z_{\text{cut}}}^{1/2} dz P_{i \to jk}(z) F_k(x - f(z), \mu) \right)
\]

where we have simplified using the identity \( P_{i \to jk}(z) = P_{i \to kj}(1 - z) \). This evolution equation is linear, and hence is numerically no more difficult to solve than the ordinary DGLAP equations. This form holds both for the ordinary soft drop multiplicity as well as for the weighted variants in App. C.1, just with a different choice of \( f(z) \).

### 4.5.3 Evolution for Pure Yang-Mills

Before showing numerical results, it is instructive to consider the case of \( n_f = 0 \), where there is only a gluon GFF and the evolution can be studied analytically. Of course, this limit cannot teach us anything about quark/gluon discrimination directly, but we will see that the gluon GFF asymptotes to an exact Poisson distribution at sufficiently large \( \mu \), such that it behaves like an idealized counting observable.
For pure Yang-Mills, we can drop flavor labels, and write the gluon GFF as \( F \equiv F_g \) and the relevant splitting function as \( P(z) \equiv P_{g \to gg}(z) \). Specializing to soft drop multiplicity (i.e. \( f(z) = 1 \)), the evolution equation in Eq. (4.5.7) becomes

\[
\frac{\mu}{d\mu} F(x, \mu) = \frac{\alpha_s(\mu)}{\pi} \left( \int_0^{z_{\text{cut}}} dz P(z) F(x, \mu) + \int_{z_{\text{cut}}}^{1/2} dz P(z) F(x - 1, \mu) \right) \tag{4.5.8}
\]

\[
= P_{\text{ave}} \frac{\alpha_s(\mu)}{2\pi} \left( F(x - 1, \mu) - F(x, \mu) \right), \tag{4.5.9}
\]

where we have defined

\[
P_{\text{ave}} = \int_{z_{\text{cut}}}^{1-z_{\text{cut}}} dz P(z). \tag{4.5.10}
\]

The interpretation of Eq. (4.5.9) is that gluon emissions that pass soft drop are added at a rate of \( P_{\text{ave}} \alpha_s(\mu)/2\pi \) in log \( \mu \) evolution. Specifically, in evolving from \( \mu_i \) to \( \mu_f \), the expected number of additional emissions is

\[
\lambda(\mu_i, \mu_f) = P_{\text{ave}} \int_{\log \mu_i}^{\log \mu_f} d(\log \mu) \alpha_s(\mu), \tag{4.5.11}
\]

so the GFF at \( \mu_f \) is

\[
F(x, \mu_f) = F(x, \mu_i) \otimes \text{Pois}(\lambda(\mu_i, \mu_f))[x], \tag{4.5.12}
\]

where the convolution is in \( x \).\(^{12}\)

As \( \mu_f \) increases, more emissions are added, so the initial GFF distributions at \( \mu_i \) becomes less and less important. Substituting in the one-loop running of the strong

\[^{12}\text{The reader who finds this derivation too slick can explicitly check that Eq. (4.5.12) solves Eq. (4.5.9). It is helpful to note that } \frac{d}{dx} \text{Pois}(\lambda)[x] = \text{Pois}(\lambda)[x - 1] - \text{Pois}(\lambda)[x].\]
coupling constant in pure Yang-Mills,
\[
\alpha_s(\mu) = \frac{1}{\beta_0 \log(\mu^2/\Lambda_{\text{QCD}}^2)}, \quad \beta_0 = \frac{11}{3} C_A, \tag{4.5.13}
\]
the number of expected emissions is
\[
\lambda(\mu_i, \mu_f) = \frac{P_{\text{ave}}}{4\pi \beta_0} \log\left(\frac{\log \frac{\mu_f}{\Lambda_{\text{QCD}}}}{\log \frac{\mu_i}{\Lambda_{\text{QCD}}}}\right). \tag{4.5.14}
\]
Since this quantity continues to grow at high \(\mu_f\), the IR boundary condition \(\mathcal{F}(x, \mu_i)\) is irrelevant in the \(\mu_f \to \infty\) limit, yielding the asymptotic form
\[
\mathcal{F}(x, \mu \gg \Lambda_{\text{QCD}}) \approx \text{Pois}(\lambda(\mu))[x], \quad \lambda(\mu) = \frac{P_{\text{ave}}}{4\pi \beta_0} \log \log \frac{\mu}{\Lambda_{\text{QCD}}}. \tag{4.5.15}
\]
Thus, we find a Poisson distribution whose mean scales as \(\log \log \mu\), such that the soft drop multiplicity acts like an idealized counting observable.

### 4.5.4 Comparison to Parton Showers

We now compare the results of the GFF approach to parton shower predictions. First, in Fig. 4.15, we show the predicted discrimination power for the collinear-unsafe \(n_{SD}\) from the same four parton showers studied in Sec. 4.4.3. We see that for low \(z_{\text{cut}}\) values, the discrimination power of the collinear-unsafe soft drop multiplicity approaches that of our benchmark IRC-safe soft drop multiplicity, previously shown in Fig. 4.13. (It does not, however, reach the power of the nonperturbative soft drop multiplicities shown in Fig. 4.10a.) Making \(z_{\text{cut}}\) any smaller does not significantly improve discrimination power, so we use \(z_{\text{cut}} = 0.02\) as our baseline parameter choice.
Figure 4.15: Same as Fig. 4.13, but for the collinear-unsafe soft drop multiplicity with $\theta_{\text{cut}} = 0$ and $\beta = 0$. 
To make a prediction using the GFF approach, we need to extract the nonperturbative distributions at a low scale and then evolve them to a higher scale. In a full analysis, the low scale distributions would be extracted from data, but here we can use the parton shower generators. For this, we switch to $e^+e^-$ collisions, generating pure quark and gluon samples through the processes $e^+e^- \rightarrow \gamma/Z^* \rightarrow q\bar{q}$ and $e^+e^- \rightarrow H^* \rightarrow gg$ in VINCIA 2.0.01. Setting $R_0 = 0.6$ as our baseline, we generate jets with energies in a 10% window of $E_{\text{jet}} = 400\,\text{GeV}$, corresponding to $\mu = E_{\text{jet}}R_0 = 240\,\text{GeV}$. We then extract $n_{\text{SP}}$ from the generated events, which at leading order, is a direct measure of the corresponding GFFs.\textsuperscript{13}

Using Eq. (4.5.7), we evolve the GFFs to 4 TeV using the energy scale in Eq. (4.5.4) and the two-loop running of $\alpha_s$.\textsuperscript{14} This evolution includes all 10 active quark and antiquark flavors, as $n_f = 5$ in this energy range.\textsuperscript{15} There are various sources of theoretical uncertainties in the evolved result, and we highlight two of them in this study. The first contribution is due to the fact that the energy scale Eq. (4.5.4) only depends on the product $E_{\text{jet}}R_0$, though the initial distributions could be extracted with any $R_0$. To estimate this uncertainty, which serves as a consistency check of the choice

\textsuperscript{13}At higher orders, one has to perform a matching calculation; see further discussion in Ref. [165].

\textsuperscript{14}Since we only consider the leading-order evolution of the GFFs, strictly speaking, only leading-order evolution of $\alpha_s$ is needed at this order. Switching to one-loop running has a negligible effect on the results of this section.

\textsuperscript{15}For simplicity, we ignore effects due to the $q \rightarrow t\bar{t}$ splitting, which would require a matching calculation to the top quark electroweak decay.
Figure 4.16: RG evolution of the collinear-unsafe soft drop multiplicity for (left column) the quark singlet GFF and (right column) the gluon GFF. Shown are the results for (top row) \( z_{\text{cut}} = 0.02 \) and (bottom row) \( z_{\text{cut}} = 0.1 \), taking distributions extracted from VINCIA at a low scale and evolving them to a higher scale. The uncertainties in the evolved distributions come from varying the jet radius used for GFF extraction and the \( \mu \) scale for the RG evolution.
Figure 4.17: RG evolution of ROC curve (quark singlet vs. gluon) for the collinear-unsafe soft drop multiplicity with (a) $z_{\text{cut}} = 0.02$ and (b) $z_{\text{cut}} = 0.1$. In both cases, there is very little evolution in the discrimination power with energy scale.

of $\mu$ scale, we also extract GFFs with $R_0 = 0.3$ and $R_0 = 0.9$, keeping $\mu$ fixed. The second contribution is from uncertainty in the absolute value of the energy scale itself. To address this, we perform evolution with both half and double the energy scale of Eq. (4.5.4). We plot the envelope of these 9 results in a shaded uncertainty band. Of course, this is only a subset of the possible GFF uncertainties, but a full study is beyond the scope of this work.

The results for $z_{\text{cut}} = 0.02$ and $z_{\text{cut}} = 0.1$ are shown in Fig. 4.16, comparing the RG-evolved results to Vincia distributions extracted at the high scale. To show a single
curve for quark jets, we plot the quark-singlet distribution

\[ Q(x, \mu) = \frac{1}{2n_f} \sum_{i \in \{u, \bar{u}, \ldots, b, \bar{b}\}} F_i(x, \mu) \]

(4.5.16)
as defined in Ref. [165] (where it is instead denoted by \( S \)). We find reasonable agreement between the RG evolution and VINCIA for both \( z_{\text{cut}} \) values, with a larger range of evolution for the case of \( z_{\text{cut}} = 0.02 \). The uncertainties in the RG evolution do not fully cover the high-scale VINCIA distribution, though it is worth emphasizing that we are only using the LO evolution equations.

In Fig. 4.17, we show the RG evolution of the quark/gluon ROC curves. Despite the fact that the \( n_{\text{SD}} \) distributions themselves exhibit significant RG evolution, the corresponding ROC curves do not change significantly with the energy scale \( \mu \). This is a key prediction of the GFF approach, and one that we can better understand by studying the moments of the GFF distributions.

### 4.5.5 Moment Space Evolution

To understand the slow evolution of the quark/gluon discrimination power, consider the evolution in moment space. Following Ref. [165], the \( n^{\text{th}} \) moment of a GFF is defined as

\[ \mathcal{F}_i(n, \mu) = \int dx x^n F_i(x, \mu). \]

(4.5.17)

In moment space, we denote the gluon GFF by \( \mathcal{G}(n, \mu) \), and the quark-singlet GFF (as defined in Eq. (4.5.16)) by \( \mathcal{Q}(n, \mu) \). To derive the moment space evolution equa-
tions, we integrate both sides of Eq. (4.5.8) against \( x^n \), shift the final integral by \( x \to x + 1 \), and then simplify the \( n^{th} \) moments with the splitting function identities

\[
\int_0^1 dz \left[ P_{g\to gg}(z) + 2n_f P_{g\to q\bar{q}}(z) \right] = 0, \quad \int_0^1 dz \, P_{q\to q\bar{q}}(z) = 0.
\]  

(4.5.18)

After these manipulations, the moment evolution equation for the \( n^{th} \) gluon or quark-singlet GFF can be written solely in terms of the difference \( \bar{G}(n) - \bar{Q}(n) \), along with lower moments \( \bar{G}(k), \bar{Q}(k) \) for \( k < n \).

For \( n = 1 \), the evolution equation for the means is

\[
\mu \frac{d}{d\mu} \left( \frac{G(1)}{Q(1)} \right) = \frac{\alpha_s}{\pi} \left[ \left( G(1) - Q(1) \right) \left( \bar{P}_{g\to gg}^{0.1/2} + 2n_f \bar{P}_{g\to q\bar{q}}^{0.1/2} \right) + \left( \bar{P}_{q\to q\bar{q}}^{0.1/2} + \bar{P}_{q\to q\bar{q}}^{0.1/2} \right) \right],
\]

(4.5.19)

where we are suppressing the \( \mu \) arguments and using the abbreviated notation

\[
\bar{P}_{i\to jk}^{z_1,z_2} = \int_{z_1}^{z_2} dz \, P_{i\to jk}(z).
\]

(4.5.20)

The appearance of the difference of the moments on the right-hand side has a dramatic effect on the high-energy limit of the evolution. Specifically, the difference in the means evolves as

\[
\mu \frac{d}{d\mu} \left( G(1) - Q(1) \right) = \frac{\alpha_s}{\pi} \left[ c_1 - c_2 \left( G(1) - Q(1) \right) \right],
\]

(4.5.21)

where \( c_1 \) and \( c_2 \) are positive constants defined by integrals of the splitting functions.

Thus, at high energies, the difference in the means asymptotes to a constant,

\[
G(1) - Q(1) \to \frac{c_1}{c_2} = \frac{P_{g\to gg}^{0.1/2} + 2n_f P_{g\to q\bar{q}}^{0.1/2} - P_{q\to q\bar{q}}^{0.1/2} - P_{q\to q\bar{q}}^{0.1/2}}{P_{g\to gg}^{0.1/2} - P_{g\to gg}^{0.1/2}}.
\]

(4.5.22)
This asymptotic behavior is strikingly different from the LL analysis of IRC-safe multiplicity in Sec. 4.3. In the IRC-safe case, the LL prediction is that the gluon and quark means should have a constant ratio determined by $C_A/C_F$. Here, in the collinear-unsafe case, the gluon and quark means asymptote to having a constant difference. Physically, this occurs because the RG evolution takes flavor mixing effects into account, so that at sufficiently high energies, the $n_{SD}$ distributions for quark and gluon jets become essentially the same. While we have only presented the calculation for the quark-singlet mean, it is straightforward to show that the means for each individual quark flavor behave in the same way, with differences between different quark flavors evolving to zero.

Moving to higher moments, a useful simplification occurs for the variances,

$$
\sigma_i^2 = \mathcal{F}_i(2) - \mathcal{F}_i(1)^2.
$$

In this case, the evolution of the variances only depends on the difference of the variances and the difference of the means,

$$
\frac{1}{\mu} \frac{d}{d\mu} \left( \frac{\sigma_i^2}{\sigma_j^2} \right) = \frac{\alpha_s}{\pi} \left[ \left( \frac{\tilde{P}_{g\rightarrow gg}^{0,1/2}}{\tilde{P}_{q\rightarrow gg}^{0,1/2}} \right) \left( \frac{\sigma_i^2}{\sigma_j^2} - (\overline{\mathcal{G}}(1) - \overline{\mathcal{Q}}(1))^2 \right) + \left( \tilde{P}_{g\rightarrow gg}^{z_{cut},1/2} + 2n_f \tilde{P}_{g\rightarrow q\bar{q}}^{z_{cut},1/2} \right) \left( \tilde{P}_{q\rightarrow q\bar{q}}^{z_{cut},1/2} + \tilde{P}_{q\rightarrow gg}^{z_{cut},1/2} \right) \left( \sigma_i^2 - \sigma_j^2 \right) \right].
$$

At sufficiently high energies, $\overline{\mathcal{G}}(1) - \overline{\mathcal{Q}}(1)$ approaches a constant, so the evolution equation for the variances is of the same form as the evolution equation for the means. We
Figure 4.18: (a) RG evolution of means and variances of the quark-singlet and gluon GFFs for the soft drop multiplicity with $z_{\text{cut}} = 0.02$. (b) RG evolution of the mean/variance differences, which asymptotically approach constants. Also shown is the relative width $w_{\text{rel}}$ defined in Eq. (4.5.26), which increases slowly. For comparison, quantities extracted from VINCIA at $E_{\text{jet}} = 4\,\text{TeV}$ are shown as dots.

find that, like the means, the difference of variances asymptotes to a constant,

$$\sigma_{\bar{G}}^2 - \sigma_{\bar{Q}}^2 \Rightarrow \text{const.} \quad (4.5.25)$$

Substituting our asymptotic results back into Eq. (4.5.19) and Eq. (4.5.24), we see that both the mean and variance simply grow linearly in $\alpha_s(\mu)d(\log \mu)$ at high energies, so that they become proportional in the UV limit. Therefore, even with flavor-mixing effects, the soft drop multiplicity maintains a Poisson-like distribution, with $\sigma^2 = O(\mu)$.

We can roughly estimate the discrimination power of the soft drop multiplicity us-
ing a relative width, similar to that of Eq. (4.3.13). Since Casimir scaling no longer holds, the distance between the quark-singlet and gluon distributions is no longer characterized by the means, but rather the difference in means. Moreover, in the UV limit, the standard deviations of the quark singlet and gluon distributions approach each other. Thus, the quantity

\[ w_{\text{rel}} \equiv \frac{\sqrt{\sigma^2}}{\bar{\mathcal{G}}(1) - \bar{\mathcal{Q}}(1)} \]  

(4.5.26)

characterizes the extent to which the distributions overlap, and hence measures the discrimination power of the soft drop multiplicity. We see that, as a result of flavor-mixing effects, the relative width is now expected to increase somewhat as more emissions are counted, roughly as the square root of the mean.

To verify these results, we numerically evolve the GFFs according to Eq. (4.5.7), starting from an initial condition extracted from VINCIA 2.0.01 at \( E_{\text{jet}} = 400 \text{ GeV} \) and \( R = 0.6 \). As in Fig. 4.16, we show a theoretical uncertainty band constructed from the envelope of 9 results. In Fig. 4.18a, we show the evolution of the mean and variance of the soft drop multiplicity for quark singlets and gluons. As expected from the above analysis, the mean and variance curves become parallel at sufficiently large values of \( \mu \). This is confirmed in Fig. 4.18b, which shows that the differences do indeed asymptote to constant values.

Crucially, the relative width in Fig. 4.18b remains approximately constant over a large energy range, as the increase in the standard deviation is canceled by the in-
crease in the mean difference as it approaches its asymptotic value. This explains the slow evolution of discrimination power seen in Fig. 4.17. In this way, even though these collinear-unsafe distributions cannot be predicted directly from first principles, the GFF approach gives us a valuable analytic handle on their RG evolution.

4.6 Conclusions

Quark/gluon discrimination has a long history, with many proposed discriminants [123, 155, 178, 180, 191, 222, 223, 237, 243, 255, 256, 279, 288] though relatively few analytic calculations [84, 253, 255]. Because $C_A/C_F$ is an order 1 number, distinguishing quark- from gluon-initiated jets is an intrinsically hard problem. Moreover, to gain a quantitative understanding of quark/gluon separation power, one has to account for physics effects beyond the LL approximation, including the impact of non-perturbative physics. These physics effects are modeled to differing degrees in parton shower generators, but ultimately one wants quark/gluon studies to be based on systematically-improvable analytic calculations.

In this paper, we introduced an IRC-safe counting observable which approaches the quark/gluon discrimination performance of IRC-unsafe track multiplicity. Through a LL analysis, we demystified the power of multiplicity, showing that Poisson distributions typically yield better quark/gluon separation than Sudakov distributions, even though they are both controlled by the same $C_A$ and $C_F$ Casimir factors. Specifically,
we introduced soft drop multiplicity, which depends on multiple soft gluon emissions even at LL accuracy, allowing it to outperform observables like jet mass whose value is dominated by a single gluon emission. Remarkably, there is a choice of ISD parameters where soft drop multiplicity is controlled by perturbative physics, such that its behavior can be reliably studied from first principles.

To gain a more quantitative understanding of $n_{\text{SD}}$, we introduced NLL evolution equations, which allowed us to make interesting comparisons to parton shower generators. We also studied a collinear-unsafe (but infrared-safe) version of $n_{\text{SD}}$, whose RG evolution could be studied using the formalism of GFFs. In both cases, analytic understanding was aided by the recursive structure of the observable. This motivates further studies into jet measurements performed on (groomed) clustering trees, which can depart significantly from the more commonly studied additive observables.

Ultimately, any single observable will never match the performance of multivariate jet tagging methods. This has been emphasized recently in the context of deep neural networks which exploit subtle correlations to maximize separation power [28, 43, 50, 135, 145, 156, 157, 211, 226, 237, 257]. Still, we are encouraged by observables like soft drop multiplicity which offer a balance between discrimination power and analytic tractability. Going beyond LL order where $n_{\text{SD}}$ can saturate the discrimination power (see Sec. 4.3.3), it would be interesting to study correlations between $n_{\text{SD}}$ and other IRC-safe observables like jet mass to see if there is additional information in their combination. Because the physics basis for $n_{\text{SD}}$ is so transparent, we suspect it
will be a useful benchmark for both parton shower tuning and experimental jet analyses. Because the analytic structure of $n_{SD}$ is so unique, we hope it inspires new precision calculations in QCD.
5

A Framework for Unsupervised Learning in Particle Physics
5.1 Introduction

Machine learning models based on deep neural networks have revolutionized information processing over the last decade. Such models can recognize objects in images [215, 220, 242], perform language translation [40, 313], transcribe spoken language [206], and even speak written text [306] at approaching human level. The truly revolutionary aspect of this progress is the generality of deep neural networks: a broad diversity of network architectures can be created from basic building blocks that allow for efficient calculation of gradients via back propagation, and thus efficient optimization through stochastic gradient descent [292]. These methods are arbitrarily expressive and can model extremely high dimensional data.

The architecture of a neural network should be designed to process information efficiently, from the input data all the way through to the network’s final output.

Indeed, it empirically seems to be the case that networks that process information evenly layer-by-layer perform very well. One example of this empirical result is that deep convolutional networks for image processing seem to perform sequentially more abstract operations as a function of depth [242]. Similarly, recurrent networks perform well on time series data, as their recurrent layers naturally describe step-by-step evolution in time [271].

The power and generality of deep neural networks has been leveraged across the sciences, and in particular in particle physics. The simplest architecture explored has
been the fully-connected network, which has successfully been applied in a wide variety of contexts, such as in identifying and splitting clusters from multiple particles in the pixel detector [15], in b-tagging [20], and in τ-identification [122]. In these basic applications, the neural network optimizes its use of some finite number of relevant physical observables for the task at hand.\(^1\) One drawback of such an approach is that the neural network is limited by the observables it is given. In fact, for these applications, other multivariate methods such as boosted decision trees often have comparable performance using the same inputs, but train faster and can be less sensitive to noise [5, 189].

As an alternative to feeding a neural network a set of motivated observables, one can feed it raw information. By doing so, one allows the network to take advantage of useful features that physicists have yet to discover. One way of preprocessing the raw data in a fairly unbiased way is through the use of jet images, which contain as pixel intensities the energy deposited by jet constituents in calorimeter cells [135]. Jet images invite the use of techniques from image recognition to discriminate jets of different origins. In [135], the pixel intensities in the two-dimensional jet image were combined into a vector, and a Fisher linear discriminant was then used to find a plane in the high-dimensional space that maximally separates two different jet classes. Treating a 2-dimensional jet image as an unstructured collection of pixel intensities,

\(^1\) For recent work on constructing a basis for neural network inputs, see [153, 154, 258], and see [238] for a linear approach that does not require neural network methods.
however, ignores the spatial locality of the problem, i.e. that neighboring pixels should have related intensities. Convolutional neural networks (CNNs), which boast reduced complexity by leveraging this spatially local structure, have since been adopted instead, and they generally outperform fully-connected networks due to their efficient feature detection. In the first applications of CNNs to jet images, on boosted $W$ detection [156] and quark gluon discrimination [237], it was indeed found that simple CNNs could generally outperform previous techniques. Since then, a number of studies have aimed to optimize various discrimination tasks using CNNs [7, 86, 129, 226, 235, 259].

While the two-dimensional detector image acts as a natural representation of a jet, especially from an experimental standpoint, the 4-momenta of individual jet constituents provide a more fundamental representation for the input to a neural network. One complication in transitioning from the jet image to its list of momenta is that, while the image is a fixed-size representation, the list of momenta will have different sizes for different jets. To avoid this problem, one could truncate the list of momenta in the jet to a fixed size, and zero-pad jets smaller than this size [286]. Alternatively, there are network architectures, namely recursive (RecNNs) and recurrent neural networks (RNNs), that handle variable length inputs naturally. With such methods, one also has the freedom to choose the order in which constituent momenta are fed into the network. In [257], a RecNN was used to build a fixed-size representation of the jet, and the authors explored various ways of ordering the momenta as
input to the network: by jet clustering algorithms, by transverse momentum, and randomly. The resulting representation of the jet was then fed to a fully-connected neural network for boosted $W$ tagging. RecNNs and RNNs have also been used in similar ways for quark/gluon discrimination [125], top tagging [164], and jet charge [181]. See also [6, 211] for jet flavor classification using tracks.

To date, the majority of applications of machine learning to particle physics employ supervised machine learning techniques. Supervised learning is the optimization of a model to map input to output based on labeled input-output pairs in the training data. These training examples are typically simulated by Monte Carlo generators, in which case the labels come from the underlying physical processes being generated. Most of the classification studies mentioned above employ this style of supervised learning, and similar techniques have also been utilized for regression tasks such as pileup subtraction [235]. Alternatively, training data can be organized in mixed samples, each containing different proportions of the different underlying processes. In this case, labels correspond to the mixed samples, and learning is referred to as weakly supervised. While full and weak supervision are very similar as computational techniques, the distinction is exceptionally important in particle physics, where the underlying physical processes are unobservable in real collider data. Early studies of weakly supervised learning in particle physics show very promising results: performance comparable to fully supervised methods was found both with low-dimensional inputs [136, 268] (a few physical observables) and with very high-dimensional in-
puts [236] (jet images).

With supervised learning, there is a notion of absolute accuracy: since every training example is labeled with the desired output, the network predicts this output either correctly or incorrectly. This is in contrast to unsupervised learning, where the machine learns underlying structure that is unlabeled in the training data. Without output-labeled training examples, there is no notion of absolute accuracy. Several recent studies have employed unsupervised learning techniques in particle physics. In [269], borrowing concepts from topic modelling in text documents, the authors extract observable distributions of underlying quark and gluon jets from two mixed samples. In [157, 284, 285], generative adversarial networks (GANs) are used to efficiently generate realistic jet images and calorimeter showers.

In this work, we explore another approach to unsupervised machine learning in particle physics, in which a deep neural network learns to compute the relative differential cross section of each data point under consideration, or equivalently, the probability distribution generating the data. The power of having access to the probability distribution underlying the data should not be underestimated. For example, likelihood ratios would provide optimal discriminants [278], and sampling from the probability distribution would provide completely data-driven simulations.

In this paper, we introduce a framework named JUNIPR: “Jets from UNsupervised Interpretable PRobabilistic models”. We also present a basic implementation of this framework using a deep neural network. This network directly computes the general
probability distribution underlying particle collider data using unsupervised learning.

The task of learning the probability distribution underlying collider data comes with challenges due to the complexity of the data. Some past studies have aimed to process collider information efficiently by using neural network architectures inspired by physics techniques already in use [95, 125, 164, 181, 211, 257]. In this paper, we take this idea one step further. We scaffold the neural network architecture around a leading-order description of the physics underlying the data, from first input all the way to final output. Specifically, we base the JUNIPR framework on algorithmic jet clustering trees. The tree structure is used, both in processing input information, and in decomposing the network's output. In particular, JUNIPR’s output is organized into meaningful probabilities attached to individual nodes in a jet’s clustering tree. In addition to reducing the complexity and increasing the efficiency of the corresponding neural network, this approach also forces the machine to speak a language familiar to physicists, thus enabling its users to interpret the underlying physics it has learned.

Indeed, one common downside associated with machine learning techniques in physics is that, though they provide powerful methods to accomplish the tasks learned in training, they do little to clarify the underlying physics that underpins their success. Our approach minimizes this downside.

Let us elaborate on the tree-based architecture used for JUNIPR’s implementation. In particle physics, events at colliders are dominated by the production of collimated collections of particles known as jets. The origin of jets and many of their proper-
ties can be understood through the fundamental theory of strong interactions, quantum chromodynamics (QCD). One insight from QCD is that jets have an inherently fractal structure, inherited from the approximate scale invariance of the fundamental theory. The fractal structure is made precise through the notion of factorization, which states that the dynamics in QCD stratify according to soft, collinear, and hard physics [137, 141, 143, 173, 175], with each sector being separately scale invariant. To capture this structure efficiently in JUNIPR, we use a kind of factorized architecture, with a dense network to describe local branchings (well-suited for collinear factorization), and a global RNN superstructure general enough to encode soft coherence and any factorization-violating effects.

One might naively expect this setup to require knowledge of the sequence of splittings that created the jet. Although there is a sequence of splittings in parton-shower simulations, the splittings are only a semi-classical approximation used to model the intensely complex and essentially incalculable distribution of final state particles. Real data is not labelled with any such sequence. In fact, there are many possible sequences which could produce the same event, and the cross section for the event is given by the square of the quantum mechanical sum of all such amplitudes, including effects of virtual particles. A proxy for this fictitious splitting history is a clustering history that can be constructed in a deterministic way using a jet-clustering algorithm, such as the $k_t$ algorithm [111, 168] or the Cambridge/Aachen (C/A) algorithm [163, 312]. There is no correct algorithm: each is just a different way to process
the momenta in an event. Indeed, there seems to be useful information in the multiple
different ways that the same event can be clustered [167, 225, 260]. Any of these algo-
rithms, or any algorithm at all that encodes the momenta of an event into a binary
tree, can be used to scaffold a neural network in the JUNIPR approach.

For practical purposes, JUNIPR is implemented with respect to a fixed jet clus-
tering algorithm. Without a fixed algorithm, the probability of the final-state parti-
cles constructed through $1 \to 2$ branchings would require marginalization over all
possible clustering histories — an extremely onerous computational task. In prin-

ciple, fixing the algorithm used to implement JUNIPR should be inconsequential for its
output, namely the probability distribution over final-state momenta, as these mo-
menta are independent of clustering algorithm. To reiterate, the JUNIPR approach
does not require the chosen clustering algorithm to agree with the underlying data-
generation process; this is demonstrated in Secs. 5.5.2 and 5.5.3 below. On the other
hand, the sequence of probabilities assigned to each branching in a clustering tree cer-
tainly depends on the algorithm used to define the tree. For example, the same fi-
nal probability $P = 10^{-22}$ could be reached with one clustering algorithm through
the sequence $P = 10^{-5} \cdot 10^{-6} \cdot 10^{-8} \cdot 10^{-3}$, or with another algorithm through
$P = 10^{-15} \cdot 10^{-2} \cdot 10^{-1} \cdot 10^{-4}$. The key idea is that, if an algorithm is chosen which
does correspond to a semi-classical parton shower, the resulting sequence of probabili-
ties may be understandable. This provides avenues for users to interpret what physics
the machine learns, and we expect that dissecting JUNIPR will be useful in such cases.
We will demonstrate this throughout the paper.

It is worth emphasizing one fundamental aspect of our approach for clarity. The JUNIPR framework yields a **probabilistic model**, not a generative model. The probabilistic model allows us to directly compute the probability density of an individual jet, as defined by its set of constituent particle momenta. To be precise, this is the probability density for those particular momenta to arise in an event, conditioned on the event selection criteria used to select the training data. As a complementary example of this, shower deconstruction [298, 299] provides a theory-driven approach to probabilistic modeling in particle physics, in which probabilities are calculated using QCD rather than a neural network. In contrast, a generative model would output an example jet, taking random noise as input to seed the generation process. Given a distribution of input seeds, the jets output from a generative model should follow the same distribution as the training data. While this means that the probability distribution underlying the data is internally encoded in a generative model, this underlying distribution is hidden from the user. Examples of generative models in particle physics include Monte Carlo event generators and, more recently, GANs used to generate jet images and detector simulations [157, 284, 285].

The direct access to the probability distribution that is enabled by a probabilistic model comes with several advantages. If two different probabilistic models are trained on two different samples of jets, they can be used to compute likelihood ratios that distinguish between the two samples. Likelihood ratios provide theoretically optimal
discriminants [278], which is indeed a major motivation for JUNIPR’s probabilistic approach. One can also sample from a probabilistic model in order to generate events, though generative models are better-suited for this application [157, 284, 285]. In addition, one can use a probabilistic model to reweight events generated by an imperfect simulator, so that the reweighted events properly agree with data.

In this paper, as a proof-of-concept, we use simulated $e^+e^-$ data to train a basic implementation of the JUNIPR framework described above. We have not yet attempted to optimize all of this implementation’s hyperparameters; however, we do find that a very simple architecture with no fine tuning is adequate. This is confirmed by its impressive discrimination power and its effective predictivity for a broad class of observables, but more rigorous testing is needed to determine whether this approach can provide state-of-the-art results on the most pressing physics problems.

The general probabilistic model, its motivation, and a specific neural network implementation of it are discussed in Sec. 5.2. A comprehensive discussion of training the model, including the data used and potential subtleties in extending the model are covered in Sec. 5.3. Results on discrimination, generation, and reweighting are presented in Sec. 5.4. We provide robustness tests and some conceptually interesting results related to factorization in Sec. 5.5, including the counterintuitive anti-$k_t$ shower generator. There are many ways to generalize our approach, as well as many applications that we do not fully explore in this work. We leave a discussion of some of these possible extensions to Sec. 5.6, where we conclude.
5.2 Unsupervised Learning in Jet Physics

To establish the framework clearly and generally, Sec. 5.2.1 begins by describing JUNIPR as a general probabilistic model, independent of the specific parametric form taken by the various functions it involves. From this perspective, such a probabilistic model could be implemented in many different ways. Sec. 5.2.2 then describes the particular neural network implementation of JUNIPR used in this paper, which has a simple but QCD-customized architecture and minimal hyperparameter tuning.

5.2.1 General Probabilistic Model

Consider a set of final-state 4-momenta $p_1, \ldots, p_n$ that we hereafter refer to as “the jet”. JUNIPR computes the probability density $P_{\text{jet}}({\{p_1, \ldots, p_n}\})$ of this set of momenta arising in an event, assuming the event selection criteria used to select the training data. This probability distribution is normalized so that, abstractly,

$$
\sum_{n=1}^{\infty} \int d^4p_1 \cdots d^4p_n P_{\text{jet}}({\{p_1, \ldots, p_n}\}) = 1,
$$

(5.2.1)

where the integral extends over the physical region of phase space. (In practice, in implementing JUNIPR we discretized the phase space into cells and assigned a measure of unity to each discrete cell. This results in $P_{\text{jet}}$ being a discrete cell-size-dependent probability distribution, but this choice is conceptually unimportant here.) A high-level schematic of JUNIPR is shown in Fig. 5.1, which emphasizes that the model does not attempt to learn the quantum-mechanical evolution that created the jet, but only
Figure 5.1: JUNIPR predicts the probability density $P_{\text{jet}}(\{p_1, \ldots, p_n\})$ of finding a given set of momenta $\{p_1, \ldots, p_n\}$ in a jet, conditioned on the jet selection criteria used to select the training data. No assumptions are made about the underlying quantum-mechanical processes that generated the jet.

meaningfully predicts the likelihood of its final-state momenta.

An unstructured model of the above form would ignore the fact that we know jet evolution is well-described by a semi-classical sequence of $1 \rightarrow 2$ splittings, due to factorization theorems [137, 141, 143, 173, 175]. A model that ignores factorization would be much more opaque to interpretation, and have many more parameters than needed due to its unnecessary neutrality. Thus, we propose a model that describes a given configuration of final-state momenta using sequential $1 \rightarrow 2$ splittings. Such a sequence is defined by a jet clustering algorithm, which assigns a clustering tree to any set of final-state momenta, so that a sequential decomposition of the probability distribution can be performed without loss of generality. We imagine fixing a specific algorithm to define the trees, so that there is no need to marginalize over all possible trees in computing a probability, a computation that would be intractable. While a deterministic clustering algorithm cannot directly describe the underlying quantum-mechanical parton evolution, that is not the goal for this model. With the algorithm set, the model as shown in Fig. 5.1 becomes that shown in Fig. 5.2.
Figure 5.2: With any fixed clustering algorithm, the probability distribution over final-state momenta can be decomposed into a product of distributions. Each factor in the product corresponds to a different step in the clustering tree. Subsequent factors are conditioned on the outcomes from previous steps, so this decomposition entails no loss of generality.

We will now formalize this discussion into explicit equations. For the rest of this section we assume that the clustering tree is determined by a fixed jet algorithm (e.g. any of the generalized $k_t$ algorithms \cite{97, 99}). The particular algorithm chosen is theoretically inconsequential to the model, as the same probability distribution over final states will be learned for any choice. Practically speaking, however, certain algorithms may have advantages over others. We will discuss the choice of clustering algorithm further in Secs. 5.5.2 and 5.5.3.

The application of a clustering algorithm on the jet constituents $p_1, \ldots, p_n$ defines a sequence of “intermediate states” $k_1^{(t)}, \ldots, k_t^{(t)}$. Here the superscript $t = 1, \ldots, n$ labels the intermediate state after the $(t - 1)^{th}$ branching in the tree (where counting starts at 1) and the subscript $i = 1, \ldots, n$ enumerates momenta in that state. To be explicit,

- the “initial state” consists of a single momentum: $k_1^{(1)} = p_1 + \cdots + p_n$;

- at subsequent steps \{\(k_1^{(t)}, \ldots, k_t^{(t)}\)\} is gotten from \{\(k_1^{(t-1)}, \ldots, k_{t-1}^{(t-1)}\)\} by a single momentum-conserving 1 → 2 branching;
• after the final branching, the state is the physical jet: \( \{k_1^{(n)}, \ldots, k_n^{(n)}\} = \{p_1, \ldots, p_n\} \).

In this notation, the probability of the jet (as shown in Fig. 5.2) can be written as

\[
P_{\text{jet}}(\{p_1, \ldots, p_n\}) = \left[ \prod_{t=1}^{n-1} P_t(k_1^{(t+1)}, \ldots, k_{t+1}^{(t+1)} | k_1^{(t)}, \ldots, k_t^{(t)}) \right] \times P_n(\text{end}|k_1^{(n)}, \ldots, k_n^{(n)}) \tag{5.2.2}
\]

Eq. (5.2.2) allows for a natural, sequential description of the jet. However, it obscures the factorization of QCD which predicts an approximately self-similar splitting evolution. Thus we decompose the model further, so that each \( P_t \) in Eq. (5.2.2) is described by a \( 1 \to 2 \) branching function that only indirectly receives information about the rest of the jet. The latter is achieved via an unobserved representation vector \( h^{(t)} \) of the global state of the jet at step \( t \). To be explicit, let \( k_m^{(t)} \to k_{d_1}^{(t+1)} k_{d_2}^{(t+1)} \) denote the branching of a mother into daughters that achieves the transition from \( k_1^{(t)}, \ldots, k_t^{(t)} \) to \( k_1^{(t+1)}, \ldots, k_{t+1}^{(t+1)} \) in the clustering tree. Then we can write

\[
P_t(k_1^{(t+1)}, \ldots | k_1^{(t)}, \ldots) = P_{\text{end}}(0|h^{(t)}) P_{\text{mother}}(m^{(t)}|h^{(t)}) P_{\text{branch}}(k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)} | k_m^{(t)}, h^{(t)})
\]

\[
P_n(\text{end}|k_1^{(n)}, \ldots, \ldots) = P_{\text{end}}(1|h^{(n)}) \tag{5.2.3}
\]

where \( m^{(t)} \) is the mother’s discrete index in the \( t^{th} \) intermediate state. We thus have a sequential model that at each \( t \) step predicts

• \( P_{\text{end}}(0|h^{(t)}) \): probability over binary values for whether or not the tree ends;

• \( P_{\text{mother}}(m^{(t)}|h^{(t)}) \): probability over \( m \in \{1, \ldots, t\} \) indexing candidate mother momenta;
\[
P_{\text{branch}}(k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)} | k_m^{(t)}, h^{(t)}): \text{probability over possible } k_m \rightarrow k_{d_1}, k_{d_2} \text{ branchings.}
\]

Note that we have left the conditioning on end \(= 0 \) implicit in \( P_{\text{mother}} \) and \( P_{\text{branch}} \), since we will never need to use these functions when end \( = 1 \). In the product of Eq. (5.2.3), each subsequent factor is thus conditioned on the outcomes of previous factors, so that breaking up \( P_{\text{jet}} \) in this way is without loss of generality. In particular, no assumption has been made about the underlying physical processes that generate the data.

With these choices, we force the hidden representation \( h^{(t)} \) to encode all global information about the tree, since it must predict whether the tree ends, which momentum branches next, and the branching pattern. In fact, providing \( P_{\text{branch}} \) with the momenta that directly participate in the \( 1 \rightarrow 2 \) branching means that \( h^{(t)} \) only needs to encode global information. We show that the global structure stored in \( h^{(t)} \) is crucial for the model to predict the correct branching patterns in Sec. 5.5.1.

### 5.2.2 Neural Network Implementation

For a neural network based implementation of the model defined by Eqs. (5.2.2) and (5.2.3), we use an RNN with hidden state \( h^{(t)} \) augmented by dense neural networks for each of the three probability distributions in Eq. (5.2.3). The recurrent structure of this implementation is shown in Fig. 5.3, which emphasizes how the RNN’s hidden
Figure 5.3: Information about the clustering tree is embedded in the hidden state $h^{(t)}$ of the RNN. For brevity, this recurrent structure is simplified on the right using a shaded box to indicate stepping from $t - 1$ to $t$. At each step, the next two daughter momenta emerging in the tree and the previous hidden state $h^{(t-1)}$ are inputs to the updated hidden state $h^{(t)}$.

representation $h^{(t)}$ keeps track of the global state of the jet, by sequentially reading in the momenta that branched most recently.

The fact that $h^{(t)}$ learns and remembers the full jet, despite only being shown the two new momenta at step $t$, is ensured by the tasks for which $h^{(t)}$ is responsible. These are shown in the detailed network diagram of Fig. 5.4. There one can see that $h^{(t)}$ is the only input into the components of the model that predict when the tree ends and which momentum is next to branch. The domains of the three probability functions in Eq. (5.2.3) are shown in Fig. 5.4 as well: $P_{\text{end}}$ is defined over the binary set $\mathbb{Z}_2$ corresponding to “end” or “not”; $P_{\text{mother}}$ is multinomial over the set $\mathbb{Z}_t$ of candidate mothers; and $P_{\text{branch}}$ is defined on the space of possible $1 \rightarrow 2$ branchings, which is (a subset of) $\mathbb{R}^4$ by momentum conservation. At each step, the
Figure 5.4: Neural network implementation of the general probabilistic model proposed in Eqs. (5.2.2) and (5.2.3). The network takes as external inputs two daughter momenta and one mother momentum. The global RNN then passes only its representation vector $h^{(t)}$ to each of the dense networks shown. The networks output three full probability distributions, which predict the end of the tree, the next mother to branch, and its daughter momenta.

The model outputs the full probability distributions, which in mathematical notation are $P_{\text{end}}(Z_2|h^{(t)})$, $P_{\text{mother}}(Z_t|h^{(t)})$, and $P_{\text{branch}}(\mathbb{R}^4|k^{(t)}_m, h^{(t)})$.

Fig. 5.3 and Fig. 5.4 show how JUNIPR provides a probability distribution at each step $t$ given the momenta emerging from the preceding branching. For clarity, Fig. 5.5 separately shows how JUNIPR is used to evaluate the full probability density $P_{\text{jet}}$ over final-state momenta in a jet. At each step $t$, the point in $Z_2$ representing whether the
tree ends, the point in $\mathbb{Z}_t$ representing which mother momentum branches, and the point in $\mathbb{R}^4$ representing its daughters are plugged into the probability distributions to obtain the probabilities that should be assigned to the jet under consideration. The product of these three probabilities, taken over all $t$ steps, leads to $P_{\text{jet}}$.

Let us now go into detail about the neural network architecture used. We use basic RNN cells [204] with tanh activation,

$$h^{(t)} = \tanh \left( W \cdot \left( k^{(t)}_{d_1}, k^{(t)}_{d_2} \right) + V \cdot h^{(t-1)} + b \right),$$

(5.2.4)

and found that a hidden representation vector $h^{(t)}$ of generic size 100 was sufficient for our needs. We found GRU [132] and LSTM [217] cells to be unnecessarily complex and high-capacity for the tasks carried out in this paper. This is in contrast to language modelling, for which basic RNN cells are underpowered. To see why this might heuristically be expected, note that a sentence containing 20 words is much more complex than a jet containing 20 momenta, because the words in the sentence are ordered, whereas the momenta in the jet are not. This introduces an additional factor of $20! \sim 10^{18}$ to the complexity of language modelling. It is thus reasonable to expect that jet physics will not require all the high-powered tools designed for natural language processing.

For $P_{\text{end}}$ we use a fully-connected network with $h^{(t)}$ as input, a single hidden layer of size 100 with ReLU activation, and a sigmoid output layer. We use the same setup for $P_{\text{mother}}$, the only difference being that the output layer is a softmax over the $t$ can-
Figure 5.5: Using JUNIPR to evaluate the probability density over final-state momenta in a jet. For a given jet and its particular clustering tree, the values associated with the tree ending, which momenta branch, and the emerging daughters are all known and plugged into the probability distributions directly. The probability density of the jet is then the product over the three distributions, over all splitting steps $t$. 
didate mother momenta, ordered by energy. These choices are generic and not highly tuned. We found that JUNIPR works well for a very general set of architectures and sizes, so we stick with this simple setup.

For the branching function $P_{\text{branch}}$ we must describe the probability distribution over all possible configurations of daughter momenta $k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)}$ consistent with the mother momentum $k_m^{(t)}$. For this system, we use coordinates $x = (z, \theta, \phi, \delta)$ centered around the mother, where $z$ is the energy fraction of the softer daughter, $\theta$ ($\delta$) is the opening angle of the softer (harder) daughter, and $\phi$ specifies the plane in which the branching occurs. See Fig. 5.6 for a visualization of these coordinates.

There are two separate approaches one could take to model the branching function $P_{\text{branch}}$. Firstly, the variables $x$ could be treated as discrete, with $P_{\text{branch}}$ outputting a softmax probability over discrete cells representing different $x$ values. Secondly, one could treat $x$ as a continuous variable and use an “energy model” of the form $P_{\text{branch}} \sim e^{E(x)} / Z$, where $Z$ is a normalizing partition function. In this work we
predominantly adopt the former approach, as it is much faster, and most distributions are insensitive to the discretization of $x$. However, we do train an energy model to show that models with continuous $x$ are possible, which we discuss in Sec. 5.3.4.

In the discrete case, we bin the possible values of $x$ into a 4-dimensional grid with 10 bins per dimension, so that the entire grid has $10^4$ cells. For a given value of $x$, we place a 1 in the bin corresponding to that value, and we place 0’s everywhere else. This 1-hot encoding of the possible values of $x$ allows us to use a softmax function at the top layer of the neural network describing $P_{\text{branch}}$ (see Fig. 5.4). Furthermore, we use a dense network with a single hidden layer of size 100 and ReLU activation for $P_{\text{branch}}$, just as we did for $P_{\text{end}}$ and $P_{\text{mother}}$. The hidden units in this network receive $h^{(t)}$ as input, as well as the mother momentum $k_{\text{m}}^{(t)}$.

Thus we have a neural network implementation of Eqs. (5.2.2) and (5.2.3), with a representation of the evolving global jet state stored in $h^{(t)}$, and with fully-connected networks describing $P_{\text{end}}$, $P_{\text{mother}}$, and $P_{\text{branch}}$. As defined above, the model has a single $10^6$ parameter matrix, mapping the branching function’s 100 dimensional hidden layer to its $10^4$ dimensional output layer, and has $6 \times 10^4$ parameters elsewhere. One might refer to this implementation as JUNIPR$_0$, as one can imagine many alternative implementations within the JUNIPR framework that may prove useful in future applications. We will continue to use the term JUNIPR for brevity, to refer both to the framework and to the basic implementation described here.
5.3 Training and Validation

We now describe how to train the model outlined in Sec. 5.2.2. We begin by discussing the training data used, followed by our general approach to training and validation. Finally we discuss an alternative model choice that allows higher resolution on the particle momenta.

5.3.1 Training Data

To enable proof-of-concept demonstrations of JUNIPR’s various applications, we train the implementation described in Sec. 5.2.2 using jets simulated in PYTHIA v8.226 [2, 295] and clustered using FASTJET v3.2.2 [99]. We simulated 600k hemisphere jets in PYTHIA using the process $e^+e^- \to q\bar{q}$ at a center-of-mass energy of 1 TeV, with hemispheres defined in FASTJET using the exclusive $k_t$ algorithm [111, 168], and with an energy window of 450–550 GeV imposed on the jets. To create the deterministic trees that JUNIPR requires, we reclustered the jets using the C/A clustering algorithm [163, 312], with $E_{\text{sub}} = 1$ GeV and $R_{\text{sub}} = 0.1$. The nonzero values of $E_{\text{sub}}$ and $R_{\text{sub}}$ make the input to JUNIPR formally infrared-and-collinear safe, but this is by no means necessary. Furthermore, our approach is formally independent of the reclustering algorithm chosen. We demonstrate this by showing results using an absurd reclustering algorithm inspired by a 2D printer in Sec. 5.5.2, as well as for anti-$k_t$ [97] reclustering in Sec. 5.5.3.
Thus we have 600k quark jets with $E_{\text{jet}} \sim 500$ GeV and $R_{\text{jet}} \sim \pi/2$. We use 500k of these jets for training, with 10k set aside as a test set to monitor overfitting, and we use the remaining validation set of 100k jets to make the plots in this paper.

In the applications of Sec. 5.4, we also make use of several other data sets produced according to the above specifications, with small but important changes. We list these modifications here for completeness. In one case, quark jets from $e^+e^- \rightarrow q\bar{q}$ were required to lie in a very tight mass window of 90.7–91.7 GeV. A sample of boosted $Z$ jets from $e^+e^- \rightarrow ZZ$ events was also produced with the same mass cut. And finally, another sample of quark jets was produced, as detailed above, but with the value of $\alpha_s(m_Z)$ in the final state shower changed from PYTHIA’s default value of 0.1365 to 0.11.

Before being fed to JUNIPR, jets in these data sets must be clustered, so that each jet becomes a tree of $1 \rightarrow 2$ branchings ending in the $n$ final-state momenta of the jet:

$$
\text{jet} = \begin{pmatrix}
p_1 \\
p_2 \\
p_3 \\
\vdots \\
p_n
\end{pmatrix} \xrightarrow{\text{clustering algorithm}} \begin{pmatrix}
k_1^{(1)} & k_1^{(2)} & \ldots & k_{n-1}^{(n-1)} \\
k_2^{(1)} & k_2^{(2)} & \ldots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
k_{n-1}^{(1)} & k_{n-1}^{(2)} & \ldots & p_n
\end{pmatrix}
$$

(5.3.1)

where the momenta in one column are equal to those of the next column except for a single $1 \rightarrow 2$ branching. At each step $t$, only the momenta associated with this $1 \rightarrow 2$ branching are fed into JUNIPR, as detailed in Sec. 5.2. With this setup, JUNIPR
requires minimal parameters; it learns to update $h^{(t)}$ as the tree evolves by focusing only on the step-by-step changes to the jet. Note also that jets of arbitrary length can be considered.

Note that in implementing JUNIPR, we do not directly evaluate the branching function $P_{\text{branch}}(k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)}|k_m^{(t)}, h^{(t)})$ on the momenta $k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)}$ but instead use the parameterization $x = (z, \theta, \phi, \delta)$ shown in Fig. 5.6. In fact, we use a nonlinear transformation of this parameterization:

\[
\begin{align*}
\tilde{z} &= \frac{\log z - \log \frac{E_{\text{sub}}}{E_{\text{jet}}}}{\log \frac{1}{2} - \log \frac{E_{\text{sub}}}{E_{\text{jet}}}} \\
\tilde{\theta} &= \frac{\log \theta - \log \frac{R_{\text{sub}}}{2}}{\log R_{\text{jet}} - \log \frac{R_{\text{sub}}}{2}} \\
\tilde{\phi} &= \frac{\phi}{2\pi} \\
\tilde{\delta} &= \frac{\log \delta - \log \frac{E_{\text{sub}}R_{\text{sub}}}{E_{\text{jet}}}}{\log \frac{R_{\text{jet}}}{2} - \log \frac{E_{\text{sub}}R_{\text{sub}}}{E_{\text{jet}}}}
\end{align*}
\] (5.3.2)

This invertible transformation simply maps the range of each coordinate onto $[0, 1]$, which reduces the amount of global parametric shift required in optimization. Similarly, we perform a transformation on the components of $k_{d_1}^{(t)}, k_{d_2}^{(t)}$ before feeding them into the update rule for $h^{(t)}$ in Eq. (5.2.4); we do the same for $k_m^{(t)}$, the input to the branching function $P_{\text{branch}}$. This is a technical point that is not conceptually important.

### 5.3.2 Approach to Training

To train JUNIPR, we maximize the log likelihood over the full set of training data:

\[
\log \text{likelihood} = \sum_{\text{jet } i \text{ in data}} \log P_{\text{jet}}(\{p_1^{(i)}, \ldots, p_n^{(i)}\}).
\] (5.3.3)
For a particular jet with final-state momenta \(p_1, \ldots, p_n\) we use Eqs. (5.2.2) and (5.2.3) to compute

\[
\log P_{\text{jet}}(\{p_1, \ldots, p_n\}) = \sum_{t=1}^{n-1} \left[ \log P_{\text{end}}(0|h^{(t)}) + \log P_{\text{mother}}(m^{(t)}|h^{(t)}) + \log P_{\text{branch}}(k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)}|k_m^{(t)}, h^{(t)}) \right] + \log P_{\text{end}}(1|h^{(n)})
\]

where \(m^{(t)}\) is the index of the mother momentum at step \(t\) in the training example and \(k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)}\) are its daughters. Maximizing the log likelihood in this way allows the model to learn each \(t\) step in parallel, providing computational efficiency and stability.

For all models presented in this paper, we use basic stochastic gradient descent with the following learning rate and batch size schedule, where training proceeds from left to right:

<table>
<thead>
<tr>
<th>Schedule</th>
<th>5 epochs</th>
<th>5 epochs</th>
<th>5 epochs</th>
<th>5 epochs</th>
<th>5 epochs</th>
<th>5 epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>learning rate</td>
<td>(10^{-2})</td>
<td>(10^{-3})</td>
<td>(10^{-4})</td>
<td>(10^{-3})</td>
<td>(10^{-4})</td>
<td>(10^{-5})</td>
</tr>
<tr>
<td>batch size</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

We follow such a schedule to slowly increase the resolution and decrease the stochasticity of gradient descent throughout training. Decreasing the learning rate reduces the step size, thereby allowing finer details of the cost surface to be resolved. Increasing the batch size reduces the stochasticity by improving the sample estimates of the true gradients.
We wrote JUNIPR in Theano [27] and trained it on 16-core CPU servers using the SherlockML technical data science platform. Training JUNIPR on 500k jets according to the above schedule took an average of 4 days.

5.3.3 Validation of Model Components

JUNIPR is constructed as a probabilistic model for jet physics by expanding $P_{\text{jet}}$ as a product over steps $t$ in the jet’s clustering tree, as shown in Eq. (5.2.2). Each step involves three components: the probability $P_{\text{end}}$ that the tree will end, the probability $P_{\text{mother}}$ that a given momentum will be the next mother to branch, and the probability $P_{\text{branch}}$ over the daughter momenta of the branching, as shown in Eq. (5.2.3). We now validate each of JUNIPR’s components using our validation set of 100k previously unseen PYTHIA jets. In this section, we present histograms of actual outcomes in the PYTHIA validation set (i.e. frequency distributions) as well as JUNIPR’s probabilistic output when evaluated on the jets in this data set (i.e. marginalized probability distributions) to check for agreement.

In Fig. 5.7 we show the probability $P_{\text{end}}$ that the tree should end, as a function of both intermediate state length and maximum particle-to-jet-axis angle. In both cases we see excellent agreement with the validation data, demonstrating a good model fit with low underfitting and no overfitting. Note that Fig. 5.7 (left) is in one-to-one correspondence with the jet constituent multiplicity, and that the shape of Fig. 5.7
Figure 5.7: Validation of $P_{\text{end}}$, the probability that the tree should end. Comparison is made between actual outcomes in the validation set of PYTHIA jets and JUNIPR’s probabilistic predictions for these jets. (Left) $P_{\text{end}}$ as a function of intermediate state length. (Right) $P_{\text{end}}$ as a function of the maximum angle between the jet axis and momenta in the intermediate state.

(right) is a direct consequence of C/A clustering with $R_{\text{sub}} = 0.1$. Indeed, if an opening angle near $R_{\text{sub}}$ already exists in an angular-ordered tree, then there are likely no remaining branchings in the clustering tree.

In Fig. 5.8 we show the probability $P_{\text{mother}}$ that a given candidate will be the next mother to branch in the clustering tree, as a function of both the candidate’s index (which is sorted to be decreasing in energy) and the candidate’s angle from the jet axis. The first of these results is shown in particular for the $t = 10^{th}$ step in the clustering trees. We observe again that the model fits the validation data well. Note from Fig. 5.8 (left) that the highest energy branches of the clustering tree are most likely to undergo subsequent branchings, in line with the expectation at leading logarithmic accuracy. Fig. 5.8 (right) shows consistent predictions, since the highest energy
Figure 5.8: Validation of $P_{\text{mother}}$, the probability that a given candidate will branch next in the clustering tree. Comparison is made between actual outcomes in the validation set of PYTHIA jets and JUNIPR’s probabilistic predictions for these jets. (Left) $P_{\text{mother}}$ at $t = 10$, as a function of a candidate’s index in the energy ordered intermediate state. (Right) $P_{\text{mother}}$ averaged over all $t$’s, as a function of a candidate’s angle relative to the jet axis.

branches also lie at the narrowest angles to the jet axis.

In Fig. 5.9 we show the branching function $P_{\text{branch}}$, the component of the model that predicts how a mother momentum should split into a pair of daughter momenta. We show the branching function results for $z$ and $\theta$ (i.e. with $P_{\text{branch}}$ marginalized over the variables not shown) at the first step in the jet evolution $t = 1$, as well as at a later step $t = 10$. (See Fig. 5.6 for definitions of $z$ and $\theta$ and Eq. (5.3.2) for their ranges in the data.) This shows the dependency of the branching function on the evolving jet representation $h(t)$, which we will discuss in detail in Sec. 5.5.1. We see that for these direct predictions, JUNIPR fits the validation data almost perfectly. Note that in Fig. 5.9 (top) soft wide-angle emissions are the norm at the earliest $t$
steps, as expected with the C/A clustering algorithm. In Fig. 5.9 (bottom) one can see that later in the clustering trees, harder more-collinear branchings are commonplace. It bears repeating that these trends are highly dependent on the chosen clustering algorithm and have no precise connection to the underlying physical processes generating the data.

5.3.4 Increasing the Branching Function Resolution

In this section, we discuss increasing the resolution of the branching function

\[ P(x) = P_{\text{branch}}(k_{d_1}^{(t+1)}, k_{d_2}^{(t+1)} | k_{m}^{(t)}, h^{(t)}) \]  

including the case where \( P(x) \) is an energy model over continuous \( x = (z, \theta, \phi, \delta) \).

(The \( x \) coordinates were defined in Fig. 5.6.) This technical section can easily be skipped without loss of the logical flow of the paper.

We begin by briefly discussing increasing the resolution of the branching function over discrete \( x \), the case described in Sec. 5.2.2. The first thing to note is that with a softmax over 4-dimensional \( x \), the size of the matrix multiplication required in a dense network is quartic in the number of bins used for each dimension. We generically use 10 bins for each of \( z, \theta, \phi, \delta \) resulting in an output size of \( 10^4 \). (In fact we use 10 linearly spaced bins in the transformed coordinates of Eq. (5.3.2), and this can be seen on the logarithmic axes of Fig. 5.9, but this detail is not conceptually important.)

Given this quartic scaling, simply increasing the number of discrete \( x \) cells quickly be-
Figure 5.9: Validation of $P_{\text{branch}}$, the 4-dimensional probability distribution over $1 \rightarrow 2$ branchings. Comparison is made between actual outcomes in the validation set of \textsc{Pythia} jets and \textsc{JuniPR}'s probabilistic predictions for these jets. Results are shown for energy fraction $z$ (left) and branching angle $\theta$ (right) as defined in Fig. 5.6. Evolution step $t = 1$ is shown (top) where soft wide-angle emissions are the norm, as expected in the C/A tree. Evolution step $t = 10$ (bottom) gives rise to harder more-collinear branchings.
comes prohibitively computationally expensive. Potential solutions to this problem include: (i) using a hierarchical softmax [272, 275], and (ii) simply interpolating between the discrete bins of the model.

In a hierarchical softmax, a low-resolution probability is predicted first, say with $5^4$ cells, then another $5^4$-celled distribution is predicted inside the chosen low-resolution cell. In principle, this gives $25^4$ resolution at only twice the computational time required for $5^4$ resolution. We briefly implemented the hierarchical softmax, and preliminary tests found it to work efficiently, but perhaps with a decrease in training stability. We chose not to pursue the hierarchical softmax further in this work, primarily because we have not seen the need for resolution much higher than $10^4$ discrete $x$ cells.

Due to its ease of use, we do employ linear interpolation between the discrete bins in our baseline model with resolution $10^4$. This comes at no extra training cost, and removes most of the effects of discretization on the observable distributions generated by sampling from JUNIPR; see Sec. 5.4.2.

We now turn to the continuous version of JUNIPR in which the branching function $P(x)$ is given by an undirected energy model:

$$P(x) = \frac{e^{E(x)}}{Z}, \quad \text{where} \quad Z = \int dx \ e^{E(x)}.$$  \hfill (5.3.6)

To model $E(x)$, we again use a fully-connected network with hidden layer of size 100, as used everywhere else, except here the output layer is left to be linear. We perform
the integral over $Z$ using importance sampling:

$$Z = \int dx \frac{q(x)}{q(x)} e^{E(x)} = \left\langle \frac{e^{E(x)}}{q(x)} \right\rangle_q \approx \frac{1}{|S|} \sum_{x_s \in S} \frac{e^{E(x)}}{q(x)} = \hat{Z}(S)$$  \hspace{1cm} (5.3.7)

where $S$ is the set of $x_s$’s sampled from the importance distribution $q$.

Unlike the discrete-$x$ version of JUNIPR, where training is relatively straightforward, the continuous-$x$ version requires a non-standard technique in training the branching function $P(x)$. This is because, although Eq. (5.3.7) provides an unbiased approximation to $Z$,

$$\langle \hat{Z} \rangle_{S \sim q} = Z,$$  \hspace{1cm} (5.3.8)

this leads to a biased estimate of the log likelihood, since

$$\langle \log \hat{Z} \rangle_{S \sim q} < \log \langle \hat{Z} \rangle_{S \sim q} = \log Z$$  \hspace{1cm} (5.3.9)

by Jensen’s inequality. Thus, every gradient step taken is systematically different from the true gradient, and this bias derails training, especially near convergence when the true gradient becomes small.

To overcome this problem, we start by computing the sample variance on our estimate $\hat{Z}(S)$, which is

$$\sigma(\hat{Z})^2 = \frac{1}{|S|-1} \sum_{x_s \in S} \left( \frac{e^{E(x)}}{q(x)} - \hat{Z}(S) \right)^2.$$  \hspace{1cm} (5.3.10)

Then the percent-error $\Delta$ in our biased estimate of the gradient is approximately

$$\Delta = \frac{1}{\sqrt{|S|}} \frac{\sigma(\hat{Z})}{\hat{Z}}.$$  \hspace{1cm} (5.3.11)

This error propagates into the log likelihood, causing the bias in Eq. (5.3.9). To mit-
igate this, we adopt a policy of monitoring $\Delta$ during training, and whenever $\Delta$ increases above some value $\Delta_{\text{threshold}}$ (a hyperparameter that we set to 2%) we double the sample size $|S|$ used to compute $\tilde{Z}(S)$. This slows down training considerably, but it effectively reduces the bias in our gradient estimates. Note that while generic importance sampling typically fails in higher dimensions, our branching function lives in only 4 dimensions, so this approach is robust using any reasonable importance distribution $q$. Indeed, we found that a uniform distribution over the transformed coordinates of Eq. (5.3.2) is a fine choice for $q$.

In Fig. 5.10 we show results for JUNIPR trained with the continuous branching function as described above. In this case, we can use arbitrarily high-resolution binning, as JUNIPR has learned a fully continuous probability density. Fig. 5.10 can be roughly compared to Fig. 5.9, where we were required to use 10 bins for each dimension of $x$.

To close this section, we note that in most cases, we expect the discretized branching function with 10 bins per dimension of $x$ to be sufficient, especially if one performs a linear interpolation on the output cells. This simple case is certainly faster to train and does not require the technique described here to avoid biased gradient estimates.
Figure 5.10: Branching function modelled by a deep undirected energy model over continuous variables $z, \theta, \phi, \delta$ that parameterize the branching. Shown is the marginalized distribution over $z$, averaged over all $t$ steps. Comparison is made between actual outcomes in the validation set of PYTHIA jets and JUNIPR’s probabilistic predictions for these jets.

5.4 Applications and Results

With JUNIPR trained and validated, we turn to some of the most interesting results it enables. Given a jet, JUNIPR can compute the probability density associated with the momenta inside the jet, conditioned on the criteria used to select the training data. To visualize this, we show a C/A-clustered PYTHIA jet in Fig. 5.11 with the JUNIPR-computed probability associated with each branching written near that node in the tree. Note that these are small discretized probabilities due to the discretized implementation of JUNIPR’s branching function described in Sec. 5.2. This is shown primarily to conceptualize the model, which is constructed to be quite interpretable as it is broken down to compute the probability of each step in the clustering history of a jet.

A direct and powerful application of the JUNIPR framework, enabled by having
access to separate probabilistic models of different data sources, is in discrimination based on likelihood ratios. We discuss discrimination in Sec. 5.4.1, along with a highly intuitive way of visualizing it. In contrast, an instinctive but indirect use of JUNIPR as a probabilistic model is in sampling new jets from it. We discuss the observable distributions generated through sampling in Sec. 5.4.2. However, sampling from a probabilistic model is often inefficient (e.g. slower than PYTHIA) compared to evaluating probabilities of jets directly. In Sec. 5.4.3 we discuss reweighting samples from one simulator to match those of another distribution. In principle, this could be used to tweak PYTHIA samples to match observed collider data simply by reweighting.

Figure 5.11: JUNIPR-computed probability assigned to example PYTHIA jet and sequentially decomposed along its C/A clustering tree. Nodes are labeled with log_{10} P_t, where P_t = P_{end} \cdot P_{mother} \cdot P_{branch} includes the product of all three components of the probability at step t, as shown in Eq. (5.2.3). Color corresponds to energy and opening angle corresponds to 3-dimensional branching angle. Probabilities are small and discrete due to the discretized branching function used in JUNIPR’s implementation.
5.4.1 Likelihood Ratio Discrimination

We expect that one of the most exciting applications of JUNIPR will be in discriminating the underlying physics that could have created a jet. For example, suppose we had two sets of jets, one set corresponding to decays of a boosted Z boson, the other set simply high-energy quarks. We could then train one copy of JUNIPR on just the boosted Z sample, giving the probability distribution $P_Z$, and another copy of JUNIPR on just the quark jets, giving $P_q$. Finally, for any new jet we could determine whether the jet was initiated by a boosted Z or by a high-energy quark by looking at the likelihood ratio:

$$\frac{P_Z(\text{jet})}{P_q(\text{jet})} > \text{threshold} \implies \text{jet is boosted Z}$$

(5.4.1)

where the threshold is set according to the location on the ROC (receiver operating characteristic) curve desired for the discrimination task at hand. In contrast to approaches that try to compute likelihood ratios like this using QCD [298, 299], the JUNIPR approach can learn the separate probability distributions directly from samples of training data.

Discrimination based on the likelihood ratio theoretically provides the most statistically powerful discriminant between two hypotheses [278]. Moreover, our setup takes into account all the momenta that define a specific type of jet. Note also that for the task of pairwise discrimination between N jet types, this unsupervised approach re-

---

2We thank Kyle Cranmer for an early discussion on this topic.
quires training $N$ probabilistic models, whereas a supervised learning approach would require training $N(N-1)/2$ classifiers. Thus, we expect likelihood-ratio discrimination using JUNIPR to provide a powerful tool.

We note further that we do not even require pure samples of the two underlying processes between which we would like to discriminate [268]. Thus, it would be feasible to discriminate based solely on real collider data. In our $Z$/quark example above, we would simply train one copy of JUNIPR on a sample of predominantly boosted-$Z$ jets, and train another copy on predominantly quark jets, and the likelihood ratio of those two models would still be theoretically optimal for $Z$/quark discrimination.

In order to get a first look at the potential of likelihood-ratio discrimination using JUNIPR, we continue with the $Z$/quark example discussed above. We use PYTHIA to simulate $e^+e^- \rightarrow q\bar{q}$ and $e^+e^- \rightarrow ZZ$ events at a center-of-mass energy of 1 TeV. We impose a very tight mass window, $90.7 - 91.7$ GeV, on the jets in each data set, so that no discrimination power can be gleaned from the jet mass. More details on the generation of the data sets were given in Sec. 5.3.1. We admit that a more compelling example of discrimination power would be for quark and gluon jets at hadron colliders, but we leave a proper treatment of that important case to future work. The toy scenario studied here serves both to prove that the probabilities output by JUNIPR are meaningful, and that likelihood ratio discrimination using unsupervised probabilistic models is a promising application of the JUNIPR framework.

In Fig. 5.12 we show the $Z$/quark separation power achieved by JUNIPR, both in
Figure 5.12: (Left) Likelihood ratio $P_Z(\text{jet})/P_q(\text{jet})$ evaluated on Pythia jets in the validation set. (Right) ROC curve for discrimination based on JUNIPR’s likelihood ratio, in comparison to the empirical 2D distribution using 2-subjettiness and constituent multiplicity. All jets used in this study have masses between 90.7 and 91.7 GeV.

terms of full likelihood ratio distributions for validation sets of $Z$ and quark jets, as well as the resulting ROC curve. For comparison, in Fig. 5.12 we also show the ROC curve achieved using a 2D likelihood ratio discriminant based on 2-subjettiness [305] and multiplicity. JUNIPR’s likelihood-ratio discrimination is clearly superior to that based on combining the most natural observables: 2-subjettiness, multiplicity, (and keep in mind the tight mass cut). Of course, these observables do not provide state-of-the-art discrimination power even in this toy scenario, but we include the comparison in this proof-of-concept to provide a sense of scale on the plot.

By design, JUNIPR naturally processes the information in jets via a recurrent mechanism that tracks the evolution of their clustering trees, and this allows users to peer inside at this structure and access the probabilities at each branching. In particular,
Figure 5.13: JUNIPR trees for visualization of discrimination power at individual nodes in the clustering history. Each node is labeled with the component of $\log_{10} P_Z(jet)/P_q(jet)$ associated with that $t$ step. Colors represent energies, and opening angles represent physical 3-dimensional branching angles. The top figure is a quark jet generated using PYTHIA, with mass between between 90.7 and 91.7 GeV; the bottom figure is a boosted-Z jet. The role that the energy distribution, opening angles, multiplicity, and branching pattern play in high-performance discrimination can be understood from such pictures.

we can consider the likelihood ratio at each step in the clustering trees to understand which branchings give rise to the greatest discrimination power. We show this in Fig. 5.13, where it is clear that JUNIPR can extract useful discriminatory information at most branchings.

Indeed, visualizing jets as in Fig. 5.13 can provide a number of insights. Unsurprisingly, we see for the quark jet (on the top) that the likelihood ratio of the first branching is rather extreme, at $10^{-3.7}$, since it is unlike the energy-balanced first
branching associated with boosted-$Z$ jets. However, we also see that almost all sub-
sequent branchings are also unlike those expected in boosted-$Z$ jets, and they combine
to provide comparable discrimination power to the first branching alone. Many ef-
fects probably contribute to this separation power at later branchings, including that
quark jets often gain their mass throughout their evolution instead of solely at the
first branching, and that the quark jet is color-connected to other objects in the global
event. Such effects have proven to be useful for discrimination in other contexts [126].

Similarly, considering the boosted-$Z$ jet on the bottom of Fig. 5.13 shows that sig-
nificant discrimination power comes not only from the first branching, but also from
subsequent splittings, as the boosted-$Z$ jet evolves as a color-singlet $q\bar{q}$ pair. Note
the presence of the predictive secondary emissions sent from one quark-subjet toward
the other. This is reminiscent of the pull observable, which has proven useful for dis-
crimination in other contexts [190]. More generally, the importance of the energy dis-
tribution, opening angles, multiplicity, and branching pattern in high-performance
discrimination can be understood from such pictures.

We are very excited by the prospect of visualizing JUNIPR’s discrimination power
on jets, based on the likelihood ratio it assigns at each branching in their clustering
trees, as in Fig. 5.13. Such visualizations could provide intuition that leads to the
development of new, human-interpretable, perhaps calculable observables for discrimi-
nation in important contexts.

We would like to make one side note about discrimination, before moving on to the
next application of JUNIPR. The statement that likelihood-ratio discrimination is optimal of course only applies in the limit of perfect models. Since this limit is never fully realized, one may worry that discrimination with JUNIPR may in fact be suboptimal. Since the two probabilistic models we use for discrimination are each trained individually to replicate a certain type of jet, they are not conditioned to focus on the differences between the two jet types, which may be very subtle in the case of a difficult discrimination task. In the realistic case of slightly imperfect models, it may be advantageous for discrimination purposes to instead train the two models to focus on the differences. To be specific, one could train the two models on the two data sets simultaneously, with the goal being to maximize the likelihood ratio on one data set and minimize it on the other. Following this method in the particular example of $Z$/quark discrimination used above, one would train the $P_Z$ and $P_q$ models on data sets $D_Z$ and $D_q$ to maximize the following quantity:

$$\sum_{jet \in D_Z} \log \frac{P_Z(jet)}{P_q(jet)} - \sum_{jet \in D_q} \log \frac{P_Z(jet)}{P_q(jet)}. \tag{5.4.2}$$

Compare this to the approach we have taken above, namely training $P_Z$ and $P_q$ to separately maximize the log likelihood of Eq. (5.3.3) on their corresponding sets of training data. This alternative training method would correspond to optimizing JUNIPR for the application of discrimination, leaving intact our ability to visualize discrimination power in clustering trees, but sacrificing the probabilistic interpretation of the model’s output. We have not tested training with Eq. (5.4.2), and thus cannot
attest to its practicality, but we suspect an approach along these lines may be useful in certain contexts.

5.4.2 Generation from JUNIPR

We now turn to a more familiar approach to jet physics, but a somewhat less appropriate usage of JUNIPR models: sampling new jets from the learned probability distribution to generate traditional observable distributions. We include this application here, not only to demonstrate this capability, but also to further validate the distribution learned by JUNIPR during unsupervised training.

Sampling from JUNIPR is relatively efficient; one simply samples from the low dimensional distributions at each step $t$ and feeds those samples forward as input to subsequent steps. In this way, one generates a full jet in many steps, as detailed in Fig. 5.14.

We used the baseline implementation of JUNIPR trained on quark jets, as described in Sec. 5.3, to generate 100k jets in this way. The resulting jet mass and constituent multiplicity distributions are plotted in Fig. 5.15 where both distributions sampled from JUNIPR match those created from our validation set of 100k PYTHIA jets withheld from training. Reasonable agreement can also be seen in the 2D distributions of Fig. 5.16.

However, there are two reasons why we do not consider JUNIPR to be built for gen-
Figure 5.14: Sampling from JUNIPR to generate jets. Draws from low-dimensional distributions at each step $t$ are fed forward to subsequent steps to ultimately generate a full jet.

eration. (These drawbacks could be avoided with a generative model; see [157, 284, 285].) The first is simply that sampling from probability distributions is generally difficult. As we just showed, it turns out that JUNIPR is relatively easy to sample from, due to its sequential structure and the fact that distributions are low-dimensional at each $t$ step. Despite this, sampling jets from JUNIPR is still much slower than generation with, for example, PYTHIA.

The second reason is more fundamental. With a sequential model structured as JUNIPR is, probability distributions at late $t$ steps in generation are highly sensitive to
Figure 5.15: Jet mass (left) and constituent multiplicity (right) distributions computed on jets sampled from JUNIPR and compared against PYTHIA jets in the validation set.

the draws made at earlier $t$ steps. Very small defects in the probability distributions at early steps cause feedback in the model that amplifies those errors. Furthermore, as a partially generated jet becomes more misrepresentative of the training data, the resulting probability distributions used at later steps are less trained, which can result in a run-away effect. All of this is to say that, for the purpose of generating jets, JUNIPR’s accuracy at early $t$ steps is disproportionately important. This is in tension with the training method undertaken in Sec. 5.3.2, namely the maximization of the log-likelihood, which prioritizes all branchings equally. Thus, we should expect that some observable distributions generated by sampling jets from JUNIPR might agree worse with the validation set of PYTHIA data than otherwise expected. We mention in passing that this second drawback could be mitigated by reweighting jets after generation, as detailed in Sec. 5.4.3 below.
Figure 5.16: 2-dimensional probability distributions with respect to jet mass and constituent multiplicity. (Left) Distribution computed using validation set of PYTHIA jets. (Right) Distribution computed using jets sampled from JUNIPR.

In fact, we have found empirically that the N-subjettiness ratio observables computed by sampling from JUNIPR do not match the held-out PYTHIA data perfectly. This can be seen in Fig. 5.17 with the 2-subjettiness distribution, where the difference between the two distributions is more significant.

We consider this disagreement to be both expected and non-diminishing of JUNIPR’s potential. Indeed, in the next section we will show how to overcome this issue, by generating samples consistent with JUNIPR’s learned probabilistic model, without ever sampling from it. In particular, the disagreement in Fig. 5.17 will be rectified in Fig. 5.18.
Figure 5.17: 2-subjettiness ratio observable computed on jets sampled from JUNIPR. Disagreement with the distribution on PYTHIA jets, due to the feedback involved in sampling from JUNIPR, is visible. This disagreement is amended in Fig. 5.18.

5.4.3 Reweighting Monte Carlo Events

Another application of the JUNIPR framework is to reweight events. For example, suppose we trained JUNIPR on data from the Large Hadron Collider (LHC) to yield a probabilistic model \( P_{\text{LHC}} \). Then one could generate a sample of new events using a relatively accurate Monte Carlo simulator, train another instance of JUNIPR on that sample to yield \( P_{\text{sim}} \), and finally reweight the simulated events by \( P_{\text{LHC}}/P_{\text{sim}} \) evaluated on an event-by-event basis. This process yields a sample of events that is theoretically equivalent to the LHC data used in training \( P_{\text{LHC}} \). The advantage of such an approach is that JUNIPR can correct the simulated events on different levels, for example using the data reclustered in \( R_{\text{sub}} = 0.1 \) subjets as we have done in this paper. However, the full simulated event has the complete hadron distributions and can
thereby be interfaced with a detector simulation. This is in many ways a simpler approach than trying to improve the simulation directly through the dark art of Monte-Carlo tuning.

This reweighting is identical to importance sampling from a proposal distribution given by the simulated data distribution \( P_{\text{sim}} \). For example, suppose one wanted to measure the distribution of an observable \( O(\text{jet}) \) at the LHC, which is given by

\[
P(O) = \int d[\text{jet}] \, P_{\text{LHC}}(\text{jet}) \, \delta(O - O(\text{jet}))
\]

\[
\approx \frac{1}{N} \sum_{\text{jet} \sim P_{\text{LHC}}} \delta(O - O(\text{jet}))
\]

where the last approximation is associated with collecting a finite amount \( N \) of LHC data in order to measure the distribution. (The reader can substitute discretized delta functions appropriate for histogramming if averse to the singular notation used in these equations.) Instead of using real data, if say a public version of \( P_{\text{LHC}} \) were available, then anyone could calculate this observable distribution using only simulated data sampled from \( P_{\text{sim}} \) as follows:

\[
P(O) = \int d[\text{jet}] \, P_{\text{sim}}(\text{jet}) \, \delta(O - O(\text{jet})) \frac{P_{\text{LHC}}(\text{jet})}{P_{\text{sim}}(\text{jet})}
\]

\[
\approx \frac{1}{N} \sum_{\text{jet} \sim P_{\text{sim}}} \delta(O - O(\text{jet})) \frac{P_{\text{LHC}}(\text{jet})}{P_{\text{sim}}(\text{jet})}.
\]

In this way, one could efficiently obtain samples of arbitrary size from \( P_{\text{LHC}} \) by reweighting samples generated by an efficient simulator. The only limitation to this process is that the simulated data must be similar to the actual target data, so that they have overlapping regions of support (formal requirement) and the weights are
not too far from unity (efficiency requirement).

As with the likelihood-ratio discrimination in Sec. 5.4.1, here we will show results in a toy scenario as a proof-of-principle. Ideally a model trained on LHC data, with all related complications, would be used to reweight Monte Carlo jets to make the simulated data indiscernible from LHC data; we leave a proper study of this to future work.

Instead, here we use two samples of jets generated using two different versions of PYTHIA. We reweight jets from one of the samples and demonstrate their agreement with the other sample. In particular, we use our baseline JUNIPR model trained on PYTHIA-generated quark jets as our “true distribution”. For the moment, we will refer to this model as $P_{\alpha_s=0.1365}$, since its training data was generated using PYTHIA’s default value of $\alpha_s(m_Z) = 0.1365$ in the final state shower. As our “simulated distribution” we will use $P_{\alpha_s=0.11}$, which was trained on quark jets generated with coupling parameter changed to $\alpha_s(m_Z) = 0.11$ in PYTHIA’s final-state shower. (See Sec. 5.3.1 for a more in-depth description of the training data used.) Our goal is to show that reweighting jets from the “simulated distribution” according to the likelihood ratio $P_{\alpha_s=0.1365}/P_{\alpha_s=0.11}$ leads to observables in agreement with the “true distribution”.

In Fig. 5.18 we demonstrate that this is indeed the case. We check this for both the 2-subjettiness and 3-subjettiness ratio observables, as well as the jet shape observable. On the left side of Fig. 5.18, one can see that in all cases, the $\alpha_s = 0.11$ distribution is clearly different from the $\alpha_s = 0.1365$ distribution. On the right side of Fig. 5.18,
Figure 5.18: (Left) Disagreement in observable distributions for two PYTHIA tunes of $\alpha_s$. Observables are the 2-subjettiness and 3-subjettiness ratio observables and the jet shape, from top to bottom. (Right) Upon reweighting the $\alpha_s = 0.11$ jets by the ratio $P_{\alpha_s=0.1365}/P_{\alpha_s=0.11}$ of learned underlying probability distributions, observable distributions exhibit good agreement.
one finds that the two distributions come into relatively good agreement once the \( \alpha_s = 0.11 \) jets are reweighted by \( P_{\alpha_s=0.1365}/P_{\alpha_s=0.11} \). This also provides further confirmation that JUNIPR learns subtle correlations between constituent momenta inside jets.

Note that it was the 2-subjettiness ratio observable that JUNIPR struggled to predict well through direct sampling (see Fig. 5.17), whereas when reweighting another set of samples, JUNIPR matches the data well on this observable (see top-right of Fig. 5.18). This corroborates the discussion in Sec. 5.4.2 concerning the difficulties in sampling directly from JUNIPR.

Before closing this section, let us reiterate one point mentioned above. For the procedure of reweighting events to be practical, the weights used should not be radically different from unity, meaning that the two distributions generating the two samples should not be too different. If this condition is not satisfied, then away from the limit of infinite statistics, a few events with very large weights could vastly overpower the rest of the events, leading to a choppy reweighted distribution with large statistical uncertainties. To avoid this problem in the toy scenario explored in this section, we found it necessary to discard roughly 0.1\% of the jets in the \( \alpha_s = 0.11 \) sample which were outliers with \( P_{\alpha_s=0.1365}/P_{\alpha_s=0.11} > 100 \). These outliers were uncorrelated with the observables shown, and we believe they resulted from imperfections in the trained model. It is clear that much more needs to be understood about the application of reweighting, but this would perhaps be more effectively done in the context of a spe-
cific task of interest involving LHC data.

5.5 Factorization and JUNIPR

In the previous section, we showed some preliminary but very exciting results for
likelihood-ratio discrimination and for the generation of observables by reweighting
simulated jets. Both of these applications require access to an unsupervised proba-
bilistic model. Next we discuss some of the more subtle internal workings of JUNIPR,
which are intimately related to the underlying physics of factorization.

In particular, we show that the hidden representation $h^{(t)}$ indeed stores important
global information about intermediate states of jets in Sec. 5.5.1. We then discuss
the clustering-algorithm independence of JUNIPR by considering two distinct clus-
tering algorithms: a “printer” algorithm in Sec. 5.5.2, where momenta are processed
left-to-right and top-to-bottom as if by an inkjet printer; and the anti-$k_t$ algorithm
in Sec. 5.5.3, which allows us to present another counterintuitive result, the anti-$k_t$
shower generator.

5.5.1 The Encoding of Global Information

We have constructed JUNIPR so that all global information about the jet is contained
in the RNN’s hidden state $h^{(t)}$. Only the branching function $P_{\text{branch}}$ receives the local
$1 \rightarrow 2$ branching information in addition to $h^{(t)}$. This forces $h^{(t)}$ to contain all the
information needed to predict when the shower should end, $P_{\text{end}}$, to predict which
momentum should branch next, $P_{\text{mother}}$, and to inform the branching function $P_{\text{branch}}$
of the relevant global structure. As the primary feature vector for all three of these
distinct tasks, $h(t)$ must learn an effective representation of the jet at evolution step $t$.

To explicitly show that $h(t)$ stores important global information about the inter-
mediate jet state at step $t$, we train a new model on our baseline quark jet data (see
Sec. 5.3.1) with the difference that we remove $h(t)$ as an input to the branching func-
tion $P_{\text{branch}}$. We expect that such a “local” branching model will not evolve correctly
as the global jet structure evolves, since all global information is being withheld. This
is indeed what we find, as can be seen in Fig. 5.19. On the left side of that figure, the
evolution of the $\theta$ distribution (defined in Fig. 5.6) from $t = 1$ to $t = 2$ is shown using
100k PyTHIA jets from our held-out set of validation data. There we see the gradual
decrease in angle as expected for C/A trees. On the right side of Fig. 5.19, the evo-
lution of the branching function is shown for the “local” branching model, and the
disagreement between this damaged model and PyTHIA is clear. Note that this pre-
diction of incorrect distributions at intermediate branchings in the C/A tree will in-
evitably lead to an incorrect probability distribution $P_{\text{jet}}(\{p_1, \ldots, p_n\})$ over final-state
momenta.

While we do not show the corresponding results from our baseline (global) model
in Fig. 5.19 to avoid clutter, the agreement with PyTHIA is essentially perfect, as one
would expect from the similar check performed in Fig. 5.9. This confirms the success
Figure 5.19: (Left) Evolution of the $\theta$ distribution from $t = 1$ to $t = 2$ in the validation set of PYTHIA jets. (Right) Corresponding evolution of the branching function as predicted by a “local” branching model without access to the hidden representation $h^{(t)}$. Disagreement between PYTHIA and this local model is clear. Not shown is the result using our baseline (global) model, which agrees perfectly with PYTHIA, as expected from Fig. 5.9.

of the jet representation $h^{(t)}$ in supplying the branching function $P_{\text{branch}}$ with important information about the global structure.

### 5.5.2 Clustering Algorithm Independence

Another subtle aspect of JUNIPR is its theoretical clustering algorithm independence. In principle, the model as described in Sec. 5.2.1 is indeed independent of the chosen algorithm, which is fixed simply to avoid a sum over all possible trees consistent with the final-state momenta. That is, for each clustering procedure chosen by the user, a different model is learned, but one that describes the same probability distribution over final-state momenta, at least formally.
However, it is not guaranteed that a given neural-network implementation of JUNIPR will work well for every clustering algorithm. We have chosen an architecture that stores the global jet physics in the RNN’s hidden state $h^{(t)}$ and the local $1 \rightarrow 2$ branching physics in the branching function $P_{\text{branch}}$. This architecture is motivated by the factorizing structure of QCD, and thus JUNIPR will most easily learn jet trees that are most similar to QCD — our primary reason for predominantly using the C/A algorithm. Consequently, though the model described in Sec. 5.2.1 is formally independent of clustering algorithm, the particular implementation adopted in Sec. 5.2.2 may weakly depend on the chosen algorithm by virtue of the ease with which it can learn the data.

To put this to the test, we have introduced a jet clustering algorithm that is nothing like QCD, but more like a 2D printer. The “printer” clustering algorithm scans the 2D jet image (i.e. the cross sectional image perpendicular to the jet axis) from right-to-left and bottom-to-top, clustering particles as it encounters them. Run in reverse (i.e. as a shower) particles are emitted from the jet core from left-to-right and top-to-bottom; this is how a jet image would be printed by an inkjet printer with a single printing tip. In Fig. 5.20 we show a single PYTHIA jet clustered using the printer algorithm. As can be seen in the jet image on the right side of Fig. 5.20, momenta are indeed emitted top-to-bottom. On the left side of Fig. 5.20, we see that any

\[3\text{We thank Eric Metodiev for this suggestion.}\]
Figure 5.20: A single PYTHIA jet clustered using the printer algorithm. Shown are its clustering tree (left) and jet image (right) in which colors correspond to energies and polar coordinates correspond to the $\theta$ and $\phi$ values of the momenta. Each momentum is labelled by its corresponding step $t$ in the clustering tree.

collinear branching structure is completely absent from the clustering tree; instead, particles are steadily emitted up-and-to-the-left.

Though JUNIPR’s neural network architecture is not optimized for the informational structure of the printer algorithm, it is still able to learn the structure, by relying much more heavily on the the jet representation $h^{(t)}$. We demonstrate this by training JUNIPR on our data set of PYTHIA-generated quark jets (see Sec. 5.3.1) clustered with the printer algorithm, thus yielding the probabilistic model $P_{\text{printer}}$. Indeed, in Fig. 5.21 one can see a jet sampled from $P_{\text{printer}}$, which correctly follows the printer structure.

As expected, however, the distributions sampled from $P_{\text{printer}}$ are not quite as good as our C/A results. On the left side of Fig. 5.22 we show the 2-dimensional distribution over jet mass and constituent multiplicity generated using 100k jets sampled di-
Figure 5.21: A single jet sampled from JUNIPR, which was trained on PYTHIA-generated quark jets that were clustered using the printer algorithm. The sampled jet emits with the correct printer structure, as can be seen by its emission tree (left) and jet image (right). Each momentum is labelled by the step $t$ at which it was emitted during generation from JUNIPR.

rectly from $P_{\text{printer}}$. Comparing to the distribution generated by PYTHIA (see the left side of Fig. 5.16) this distribution matches well. However, for the 2-subjettiness ratio observable on the right side of Fig. 5.22 we get a somewhat worse match to the PYTHIA validation data; compare this to the results of the C/A model in Fig. 5.17. Of course, we discussed in Sec. 5.4.2 why we do not expect direct sampling from JUNIPR to be perfectly reliable (and we discussed a way around this in Sec. 5.4.3), but it is still clear that such distributions are comparably worse when using the printer clustering algorithm, instead of the more natural C/A algorithm.
Figure 5.22: (Left) 2-dimensional distribution with respect to jet mass and constituent multiplicity, calculated by sampling jets directly from $P_{\text{printer}}$, an instance of JUNIPR trained on jets clustered with the printer algorithm. (Right) 2-subjettiness ratio observable distribution generated using $P_{\text{printer}}$ and compared to the corresponding distribution on PYTHIA jets in the validation set.

5.5.3 Anti-$k_t$ Shower Generator

Reassured by the results of the previous section, we next consider JUNIPR trained on PYTHIA jets reclustered with anti-$k_t$ [97]. Like the printer algorithm, anti-$k_t$ does not approximate the natural collinear structure of QCD. Unlike the printer algorithm, however, anti-$k_t$ is a very commonly used tool. For the latter reason we explore anti-$k_t$ jets here.

Perhaps the most interesting result associated with an anti-$k_t$ version of JUNIPR is that it provides access to an anti-$k_t$ shower generator. Generating an anti-$k_t$ shower is counterintuitive, because the anti-$k_t$ algorithm generally clusters soft emissions one-by-one with the hard jet core. Thus, a generator must remember where previous emissions landed in order to send subsequent emissions nearby. This is required
to reproduce the correct collinear structure in the distribution of final-state of momenta. Said in another way, since the collinear factorization of QCD is not built into the anti-\(k_t\) clustering algorithm, a local (or factorized) anti-\(k_t\) generator could not produce emissions with the correct collinear distribution. Thus, we should expect that, in an anti-\(k_t\) version of JUNIPR, higher demands will be placed on the jet representation \(h^{(t)}\) to monitor all the radiation in the jet. This is certainly possible, but not the task for which our neural network architecture is optimized.

To see to what extent an anti-\(k_t\) implementation of JUNIPR relies on the global information stored in \(h^{(t)}\), we trained two models on PYTHIA-generated quark jets clustered with anti-\(k_t\) (see Sec. 5.3.1 for more details on the training data used). One model, \(P_{\text{anti}}\), has the baseline architecture outlined in Sec. 5.2. The other, \(P_{\text{anti-local}}\), is a local branching model like the one used in Sec. 5.5.1, in which the global representation \(h^{(t)}\) is withheld as input to the branching function.

In Fig. 5.23 (bottom) we show a jet sampled from \(P_{\text{anti}}\). In this case, though the tree itself does not properly guide the collinear structure of emissions, one can see that the emission directions are highly correlated with one another, demonstrating the success of the jet representation \(h^{(t)}\) in tracking the global branching pattern. In Fig. 5.23 (top) we show for comparison a jet sampled from \(P_{\text{anti-local}}\), in which the branching function does not receive \(h^{(t)}\). In the latter case, all correlation between the emission directions is lost. This shows that the global representation \(h^{(t)}\) is crucial for a successful anti-\(k_t\) branching model.
Figure 5.23: (Top) Shower sampled from an anti-$k_t$ version of JUNIPR, but one in which the global representation $h^{(t)}$ is withheld from the branching function. Correlation between emission directions is absent in this case. (Bottom) Shower sampled from an anti-$k_t$ version of JUNIPR, using the standard architecture complete with $h^{(t)}$. Strong coherence in emission directions is clearly evident.

In Fig. 5.24 we show the 2-dimensional distribution over jet mass and constituent multiplicity, as well as the 2-subjettiness distribution, generated with $P_{\text{anti}}$. One can see that the former distribution is consistent with the distribution generated by PYTHIA in Fig. 5.16. Mild disagreement between $P_{\text{anti}}$’s 2-subjettiness distribution and PYTHIA’s can be seen on the right side of Fig. 5.24. This is on par with the agreement obtained by sampling from the C/A model in Fig. 5.17.

In Sec. 5.5.1 we saw that the RNN’s hidden state $h^{(t)}$ manages the global information in JUNIPR’s neural network architecture. This is an efficient and natural way to characterize QCD-like jets, and therefore also C/A clustering trees. Though JUNIPR is formally independent of jet algorithm (i.e. in the infinite-capacity and perfect-training limit), we might expect JUNIPR’s performance to degrade somewhat when paired with
clustering algorithms that require significantly more information to be stored in $h^{(l)}$.

This was explored in Secs. 5.5.2 and 5.5.3 using two separate non-QCD-like clustering algorithms, namely the “printer” and anti-$k_t$ algorithms. Despite these clustering algorithms being unnatural choices for JUNIPR, we were able to demonstrate conceptually interesting and novel results, such as the anti-$k_t$ shower generator. This further demonstrates that JUNIPR can continue to function well, even when the clustering algorithm chosen for implementation bears little resemblance to the underlying physical processes that generate the data.
5.6 Conclusions and Outlook

In this paper, we have introduced JUNIPR as a framework for unsupervised machine learning in particle physics. The framework calls for a neural network architecture designed to efficiently describe the leading-order physics of $1 \rightarrow 2$ splittings, alongside a representation of the global jet physics. This requires the momenta in a jet to be clustered into a binary tree. The choice of clustering algorithm is not essential to JUNIPR’s performance, but choosing an algorithm that has some correspondence with an underlying physical model, such as the angular-ordered parton shower in quantum chromodynamics, gives improved performance and allows for interpretability of the network. At JUNIPR’s core is a recurrent neural network with three interconnected components. It moves along the jet’s clustering tree, evaluating the likelihood of each branching. More generally, JUNIPR is a function that acts on a set of 4-momenta in an event to compute their relative differential cross section, i.e. the probability density for this event to occur, given the event selection criteria used to select the training sample. One of the appealing features of JUNIPR is its interpretability: it provides a deconstruction of the probability density into contributions from each point in the clustering history.

There are many promising applications of JUNIPR, and we have only been able to touch on a few proof-of-concept tests in this introductory work. One exciting use case is discrimination. In contrast to supervised models which directly learn to discrimi-
nate between two samples, JUNIPR learns the features of the samples separately. It then discriminates by comparing the likelihood of a given event with respect to alternative models of the underlying physics. The resulting likelihood ratio provides theoretically optimal statistical power. As an example, we showed that JUNIPR can discriminate between boosted $Z$ bosons and quark jets (in a very tight mass window around $m_Z$) in $e^+e^-$ events when trained on the two samples separately. With JUNIPR, it is not only possible to perform powerful discrimination using unsupervised learning, but the discrimination power can be visualized over the entire clustering tree of each jet, as in Fig. 5.13. This opens new avenues for physicists to gain intuition about the physics underlying high-performance discrimination. Such studies might even inspire the construction of new calculable observables.

Another exciting potential application of JUNIPR is the reweighting of Monte Carlo events, in order to improve agreement with real collider data. A proof-of-concept of this idea was given in Fig. 5.18, where jets generated with one PYTHIA tune were reweighted to match jets generated with another. The reason this application is important is that current Monte Carlo event generators do an excellent job of simulating events on average, but are limited by the models and parameters within them. It may be easier to correct for systematic bias in event generation by a small reweighting factor appropriate for a particular data sample, rather than by trying to isolate and improve faulty components of the model. In this context, JUNIPR can be thought of as providing small but highly granular tweaks to simulations in order to improve
agreement with data.

The JUNIPR framework was used here to compute the likelihood that a given set of particle momenta will arise inside a jet. One can also imagine more general models that act on all the momenta in an entire event, including particle identification tags, or even low-level detector data. A particularly interesting direction would be to consider applying JUNIPR to heavy ion collisions, in which the medium where the jets are produced and evolve is not yet well understood. In this case, comparing the probabilities in data to those of simulation could give insights into how to improve the simulations, or more optimistically, to improve our understanding of the underlying physics.
Appendix to Chapter 2
### A.1 Partonic diboson cross sections with mass corrections

In this appendix, we give the partonic diboson cross sections at LO including all dependence on $m_W, m_Z$. This is a straightforward modification of the formulas given in the high-energy limit in eqs. (2.3.17)–(2.3.20), (2.3.24)–(2.3.25), and (2.3.28) above.

The partonic rates including all finite-mass effects can be written as

\begin{align*}
\frac{d\hat{\sigma}}{dt}(q\bar{q} \to \gamma\gamma) &= \left( \frac{1}{2} \right) \frac{\pi \alpha_2^2 s_W^4}{N_c s^2} (2 Q^4) |A_1|^2, \\
\frac{d\hat{\sigma}}{dt}(q\bar{q} \to Z\gamma) &= \frac{\pi \alpha_2^2 s_W^2 c_W^2}{N_c s^2} (L^2 Q^2 + R^2 Q^2) |A_1|^2, \\
\frac{d\hat{\sigma}}{dt}(q\bar{q} \to ZZ) &= \left( \frac{1}{2} \right) \frac{\pi \alpha_2^2 c_W^4}{N_c s^2} (L^4 + R^4) |A_1|^2, \\
\frac{d\hat{\sigma}}{dt}(q\bar{q}' \to W^\pm \gamma) &= \frac{\pi |V_{ud}|^2 \alpha_2^2 s_W^2}{N_c s^2} \left[ Y_L^2 |A_1|^2 \right] + 2 Y_L (A_1 A_3) + 4 |A_3|^2, \\
\frac{d\hat{\sigma}}{dt}(q\bar{q}' \to W^\pm Z) &= \left( \frac{1}{2} \right) \frac{\pi |V_{ud}|^2 \alpha_2^2}{N_c s^2} \left[ s_W^2 t_W^2 Y_L^2 |A_1|^2 \right] + 2 s_W^2 Y_L (A_1 A_3) + 4 c_W^2 |A_3|^2 + \frac{|A_3|^2}{2}, \\
\frac{d\hat{\sigma}}{dt}(q\bar{q} \to W^-W^+) &= \frac{\pi \alpha_2^2}{N_c s^2} \left[ \frac{|A_1|^2}{16} \right] + \frac{(A_1 A_3)}{2} + 2 |A_3|^2 + |A_\phi|^2 + 2 |A_\phi|^2,
\end{align*}

\(301\)
where $\alpha_2, L, R$ were defined in section 2.3.3.1 and $q\bar{q}'$ is $u\bar{d}$ ($d\bar{u}$) for $W^+V^0$ ($W^-V^0$).

In $d\sigma(W^-W^+)$, the upper (lower) sign holds for $u$-type ($d$-type) quarks. In these formulas, the superscripts $\ell, r$ refer to the handedness of the incoming quarks.

The partonic cross sections above are written in terms of $A_i$s, generalizations of the $a_i$s, defined as

$$|A_1|^2 = (\hat{t} \hat{u} - m_1^2 m_2^2) \left( \frac{1}{t^2} + \frac{1}{u^2} \right) + \frac{2 \hat{s} (m_1^2 + m_2^2)}{\hat{t} \hat{u}}, \quad (A.1.7)$$

$$(A_1 A_3) = \left( N_T^\ell P_s \right) (\hat{T} \hat{U}) \left( \frac{1}{u} - \frac{1}{t} \right) + \frac{1}{4} (\hat{t} \hat{u} - m_1^2 m_2^2) \left( \frac{1}{u^2} - \frac{1}{t^2} \right), \quad (A.1.8)$$

$$|A^h_3|^2 = \left( N^h_T P_s \right)^2 (\hat{T} \hat{U}) + \delta_{ht} \left[ \frac{(N_T^\ell P_s)}{4} (\hat{T} \hat{U}) \left( \frac{1}{t} + \frac{1}{u} \right) \right.$$  

$$+ \frac{1}{32} (\hat{t} \hat{u} - m_1^2 m_2^2) \left( \frac{1}{t^2} + \frac{1}{u^2} \right) - \frac{1}{16} \frac{\hat{s} (m_1^2 + m_2^2)}{\hat{t} \hat{u}} \left. \right], \quad (A.1.9)$$

$$|A^h_\phi|^2 = \left( N^h_\phi P_s \right)^2 [\hat{t} \hat{u} + 2 \hat{s} (m_1^2 + m_2^2) - m_1^2 m_2^2] , \quad (A.1.10)$$

where $m_1, m_2$ are the masses of $V_1, V_2$. Here we have abbreviated

$$(\hat{T} \hat{U}) \equiv \hat{t} \hat{u} - \hat{s} (m_1^2 + m_2^2) - m_1^2 m_2^2 \quad (A.1.11)$$
and defined $\hat{s}$-channel propagators

$$
P_s \equiv \begin{cases} 
\frac{1}{s - m_W^2} & \text{for } W^0, \\
\frac{1}{s - m_Z^2} & \text{for } W^- W^+. 
\end{cases} \tag{A.1.12}
$$

Each $P_s$ appears with a coefficient:

$$
N^\ell_T = N^r_\phi = \frac{1}{2} \quad \text{for } W^0, \tag{A.1.13}
$$

while for $W^- W^+$,

$$
N^\ell_T = |T_3| - \left|Q\right| \frac{m_Z^2 s_W^2}{s}, \tag{A.1.14}
$$

$$
N^r_T = - \left|Q\right| \frac{m_Z^2 s_W^2}{s}, \tag{A.1.15}
$$

$$
N^\ell_\phi = i_W^2 Y_L \left( \frac{1}{2} - \frac{m_Z^2 c_W^2}{s} \right) + T_3 \left( \frac{1}{2} - \frac{m_Z^2 s_W^2}{s} \right), \tag{A.1.16}
$$

$$
N^r_\phi = i_W^2 Y_R \left( \frac{1}{2} - \frac{m_Z^2 c_W^2}{s} \right). \tag{A.1.17}
$$

One can obtain these coefficients by combining the $\gamma$- and $Z$-propagators (along with their attached coupling constants) for $\hat{s}$-channel production of $w^- w^+$ or $\phi^- \phi^+$. 
A.2 On the approximations used in $gg \to VV$

estimates

In all our studies of $gg \to V_1^0 V_2^0$, we have been using MCFM’s implementation of
the dominant $O(\alpha_s^3)$ correction to this process, which in turn relies on the calculation
of ref. [80]. This calculation computes only part of the $O(\alpha_s^2)$ and $O(\alpha_s^3)$ corrections
to $pp \to V_1^0 V_2^0$ production, leaving out many terms. One may reasonably wonder
whether it represents a consistent calculation, and whether we can safely use it to
determine overall normalizations as in figure 2.11, to relate the contributions for dif-
dferent $V_1^0 V_2^0$ processes as in eq. (2.4.5), and to estimate remaining scale dependence as
in section 2.4.4.

Our methods rely upon the fact that all $O(\alpha_s^2)$ and $O(\alpha_s^3)$ terms included in ref. [80]
contain a single fermion loop at amplitude level which is squared in the matrix ele-
ment. The loop is proportional to the number of quark generations $N_g$, so these terms
are of order $N_g^2$, and are also proportional to a particular combination of $SU(2) \times U(1)$
charges that arise because the $V_1$ and $V_2$ gauge bosons must attach to the loop.

In the perturbative expansion of $pp \to V_1^0 V_2^0$, this proportionality to $N_g^2$ arises first
in the leading $gg \to V_1^0 V_2^0$ box graph, and not in any other graph at this or lower
order. Therefore, at order $O(\alpha_s^2)$, we may parametrically separate the loop graph from
all other terms. This justifies treating the scale choices in this part of the calculation

304
as separate from those in the NLO calculation.

This would still be true at \(O(\alpha_s^2)\) if all terms proportional to \(N_f^2\) were included. Only these terms can be involved in moderating the \(\mu_R\) dependence of the \(O(\alpha_s^2)\) box graph. In turn, this would justify the methods used in figure 2.11 to fix the normalization and in section 2.4.4 to estimate remaining scale dependence. Furthermore, eq. (2.4.5) would be true, because the \(N_f^2\)-dependent terms for the three \(V_1V_2^0\) processes would differ only in the \(SU(2) \times U(1)\) factors arising where the \(\gamma\) or \(Z\) bosons attach to the fermion loop or lines (up to small effects from the top quark and from \(m_Z\)).

However, ref. [80] presents the two-loop corrections to \(gg \to \gamma\gamma\) and the radiative correction \(gg \to \gamma\gamma g\), but omits a term from the interference between the tree-level and one-loop-level amplitudes for the process \(gg \to \gamma\gamma q\bar{q}\). This interference term is also of order \(N_f^2\) and needs to be included to capture the full scale dependence of the \(O(N_f^2)\) piece at \(O(\alpha_s^2)\).

We now argue that the \(gg \to V_1^0V_2^0 q\bar{q}\) processes at \(O(\alpha_s^2)\) and \(O(\alpha_s^3)\) are subleading compared to what we have included. The contribution of the tree-level rate for \(gg \to V_1^0V_2 q\bar{q}\) was studied for \(V_1V_2 = W\gamma\) and \(WZ\) in ref. [25], and for \(Z\gamma\) in ref. [26], and it turns out to be anomalously small. In particular, even for \(Z\gamma\) it turns out to be less than 30% of the \(O(\alpha_s^2)\) \(gg \to Z\gamma\) rate, which, as discussed in section 2.4.3, itself has a partial \(SU(2) \times U(1)\) cancellation in the loop. For \(\gamma\gamma\) and \(ZZ\) we expect the corresponding graph to be significantly smaller relative to the \(O(\alpha_s^2)\) rate.
Meanwhile, at $O(\alpha_s^3)$, the interference of the corresponding amplitude against its one-loop correction does not affect the scale sensitivity of the resulting calculation. No additional $\mu_R$ sensitivity appears, since the only loops that arise are all in the form of subdiagrams that are themselves finite. Simultaneously, $\mu_F$ sensitivity is already taken into account by varying the PDF scale appropriately: the tree-level $gg \to V_1^0V_2^0q\bar{q}$ process has no QCD infrared divergences in the final state, so the interference term contributes only universal divergences, which are already included in the PDF evolution.

This (along with the recent paper [106]) gives us some confidence that our method for normalizing $gg \to ZZ$ and $\gamma\gamma$ has a small relative uncertainty. The correction for $Z\gamma$, suppressed at $O(\alpha_s^2)$, may have a larger relative uncertainty, but this is still small in absolute terms. More quantitatively, in the $R_{1a}$ ratio the contribution of $gg \to V_1^0V_2^0$ in the lowest $\bar{m}_T$ bin is roughly +5% from $Z\gamma$ and −13% from $\gamma\gamma$, and this drops off rapidly with $\bar{m}_T$. Even if we took an overly conservative 30% relative uncertainty estimate for the normalization of $gg \to Z\gamma$, and ignored any correlations with $gg \to \gamma\gamma$ that would cancel in the ratio, this would translate into at most a 1.5% ($< 1\%$) uncertainty in the lowest (highest) bin in figure 2.1 for $R_{1a}$. A similar statement applies for $R_{1c}$. 
A.3 Further discussion of NNLO $K$ factors

We have assumed NNLO corrections are small compared to our $K_{\text{NLO/LO}}$ factors of order 1.5. Also, as suggested in ref. [80], we assumed that the $gg$ loop contributions give the majority (or rather, more precisely and more importantly, the largest fraction that does not cancel in ratios) of the NNLO contributions. In several cases, recently computed fully differential NNLO cross sections for the $\gamma\gamma$, $Z\gamma$, and $ZZ$ processes feature much larger corrections than what we have claimed to expect for our observables. Here we discuss how the cases where this is true are affected by one or more of the issues we discuss in section 2.4.1, causing the $K_{\text{NNLO/NLO}}$ factors to be larger than would be the case for our cuts and observable.

In figure 1 of ref. [109], which shows the fully-differential NNLO calculation of the $\gamma\gamma$ cross section, one sees $K$ factors of $K_{\text{NLO/LO}} \sim 3$ and $K_{\text{NNLO/NLO}} \sim 1.6$. Meanwhile the $gg$ box contribution was found to be $\approx 15\%$ of the total NNLO correction. These results would appear to cast doubt on our assumptions. However, their calculation uses fixed asymmetric cuts for the photon of $p_T^{\text{hard}} \geq 40$ GeV and $p_T^{\text{soft}} \geq 25$ GeV, with invariant diphoton mass $m_{\gamma\gamma}$ as the observable. As the authors point out, this allows a large NLO contribution, at any $m_{\gamma\gamma}$, from events with $p_T^{\text{soft}} \sim 25$ GeV and a large $p_T^{\text{hard}}$, which leads to large logarithms. This explains both the much larger NLO corrections and (since the $gg$ process at NNLO only occurs with LO kinematics) the smallness of the $gg$ contribution compared to the total NNLO cor-
rection. This interpretation of the origin of the large \( K \) factors is supported by the cross section presented in ref. [133], figure 2, which, using more symmetric fixed cuts of \( p_T^{\text{hard}} \geq 25 \text{ GeV} \) and \( p_T^{\text{soft}} \geq 22 \text{ GeV} \), found \( K_{\text{NNLO/NLO}} \sim 1.3 \). Recall that we used scaling asymmetric cuts that never allow \( p_T^{\text{hard}} \gg p_T^{\text{soft}} \), so we expect we have smaller quantum corrections than in either of these cases.

Large \( K \) factors are also reported for the NNLO \( Z\gamma \) cross sections in ref. [207]; see figure 5 of that paper. Here the dominant issue is not \( p_T \) cuts (for the \( Z \), at least, the calculation imposes none), but the treatment of resonances. A cut on the photon \( p_T \) of 40 GeV (combined with lepton isolation cuts) ensures that at leading order, the \( Z \) is kinematically forbidden from decaying to \( \ell^+\ell^-\gamma \). However once the \( \ell^+\ell^-\gamma \) system is allowed to recoil off a jet, such a configuration becomes kinematically accessible, resulting in a large \( K \) factor at small \( m_{\ell^+\ell^-\gamma} \). But one sees much smaller \( K \) factors either if the \( p_T \) cut on the photon is reduced, allowing \( Z \to \ell^+\ell^-\gamma \) to arise at LO, or if \( m_{\ell^+\ell^-\gamma} \) is taken much larger than \( m_Z \). In the latter case, which is more relevant for us, it is found that \( K_{\text{NNLO/NLO}} \sim 1.2 \) for their cuts; while the \( K_{\text{NLO/LO}} \) factor is not reported, we estimate from the figure that it is \( \sim 1.7 - 1.8 \).

Meanwhile \( K_{\text{NNLO/NLO}} \) factors reported for \( ZZ \) in ref. [208], for cuts matching ATLAS and CMS analyses, are about 1.15 for the 8 TeV LHC. The figure in ref. [108] for the inclusive \( ZZ \) cross section shows a \( K_{\text{NNLO/NLO}} \) factor growing from 1.13 to 1.17 between 8 and 13 TeV, over an \( K_{\text{NLO/LO}} \) factor of about 1.5.

The same figure in ref. [108] shows that the \( gg \to ZZ \) loop contribution to the
inclusive cross section is about 60% of the total NNLO contribution. Since it is evaluated at lowest order, this percentage has a large $\mu_R$ uncertainty. Our discussion in section 2.4.3 supports using a low scale in evaluating this contribution, which may increase it further. (See also the recent result of ref. [106].) Some further evidence that the $gg$ contribution is the majority, for an appropriate observable similar to ours, was given in ref. [101], figure 6, though this involved only a partial evaluation of the NNLO contribution.

We have less evidence for the relative size of the $gg \rightarrow \gamma\gamma$ loop, except that in the case of the very large $K$ factors of ref. [109] it can be as small as 15% of the full NNLO contribution. Changing the cuts as was done in ref. [133] reduced the $K_{\text{NNLO/NLO}}$ factor to 1.3 but should have no effect on the lowest-order $gg \rightarrow \gamma\gamma$ loop calculation, which has LO kinematics, so this could bring the contribution to $\sim 30\%$. If our cuts lead to even smaller NLO $K$ factors, as we suspect, this could bring the $gg \rightarrow \gamma\gamma$ loop into the majority.

For $Z\gamma$, coherent cancellations between up and down quarks in the loop cause the $gg \rightarrow Z\gamma$ loop contribution to be a few times smaller than the other two, relative to the NLO calculation. (See our figure 2.10.) Although ref. [207] reports this loop is only 6–9% of the total NNLO correction, one can see from figure 5 of that paper that this percentage increases to $\sim 20\%$ once the $Z^*\gamma$ system is well above the $Z$ pole. This is consistent with our estimate of the $Z\gamma$ contribution being quite uncertain in relative terms but of little importance in absolute terms.
In sum, we find no clear inconsistencies in the literature between our assumptions and existing calculations.
Appendix to Chapter 3
B.1 Three-Loop $\beta$-function and Cusp Anomalous Dimension

The $\beta$-function is defined to be

$$\beta(\alpha_s) = \mu \frac{\partial \alpha_s}{\partial \mu} = -2\alpha_s \sum_{n=0}^{\infty} \beta_n \left( \frac{\alpha_s}{4\pi} \right)^{n+1}. \quad (B.1.1)$$

For NNLL resummation, we need the $\beta$-function to three-loop order [246, 304]. The first three coefficients are

$$\beta_0 = \frac{11}{3} C_A - \frac{4}{3} T_{R f}, \quad (B.1.2)$$
$$\beta_1 = \frac{34}{3} C_A^2 - 4T_{R f} \left( C_F + \frac{5}{3} C_A \right), \quad (B.1.3)$$
$$\beta_2 = \frac{2857}{54} C_A^3 + T_{R f} \left( 2C_F^2 - \frac{205}{9} C_F C_A - \frac{1415}{27} C_A^2 \right) + T_{R f}^2 \left( \frac{44}{9} C_F + \frac{158}{27} C_A \right). \quad (B.1.4)$$

For NNLL resummation, we need the cusp anomalous dimension

$$\Gamma_{\text{cusp}} = \sum_{n=0}^{\infty} \Gamma_n \left( \frac{\alpha_s}{4\pi} \right)^{n+1} \quad (B.1.5)$$

to three-loop order. The first three coefficients of the cusp anomalous dimension are
\[ \Gamma_0 = 4, \]  
\[ \Gamma_1 = 4C_A \left( \frac{67}{9} - \frac{\pi^2}{3} \right) - \frac{80}{9} T_R n_f, \]  
\[ \Gamma_2 = 4C_A^2 \left( \frac{245}{6} - \frac{134\pi^2}{27} + \frac{11\pi^4}{45} + \frac{22}{3} \zeta_3 \right) + 32C_A T_R n_f \left( -\frac{209}{108} + \frac{5\pi^2}{27} - \frac{7}{3} \zeta_3 \right) + 4C_F T_R n_f \left( 16\zeta_3 - \frac{55}{3} \right) - \frac{64}{27} T_R^2 n_f^2. \]  

**B.2 Hard Function**

The hard function for dijet production in $e^+e^-$ collisions is defined by the Wilson coefficient for matching the full QCD current onto the SCET dijet operator. For $e^+e^- \rightarrow q\bar{q}$ events, the Wilson coefficient $C \left( Q^2, \mu \right)$ is

\[ \langle q\bar{q}| \bar{\psi} \Gamma \psi |0\rangle = C \left( Q^2, \mu \right) \langle q\bar{q}| Y_n^\dagger Y_{\bar{n}} \chi_{\bar{n}} |0\rangle. \]  

Here, $\bar{\psi} \Gamma \psi$ is the QCD current for the production of a $q\bar{q}$ pair from the vacuum. $\chi_{\bar{n}}$ is a quark jet operator collinear quark operator defined in the light-like direction $\bar{n}$ in SCET. For calculations at leading power, $\chi_n = W_t^\dagger \psi$, with $W_t$ a Wilson line pointing in some direction $t$ not collinear to $n$ and $\psi$ is an ordinary quark field. The soft Wilson lines $Y_n$ and $Y_{\bar{n}}$ point in the $n$ and $\bar{n}$ directions respectively. $\Gamma$ represents a
generic Dirac matrix. We have ignored contraction with the leptonic tensor for simplicity. The Wilson lines $Y_n$ is defined as

$$Y_n(x^\mu) = \mathcal{P} \exp \left(i g \int_0^\infty ds \ n \cdot A(x^\mu + sn^\mu) \right), \quad (B.2.2)$$

where $\mathcal{P}$ denotes path-ordering. $Y_n$ and $W_t$ are defined similarly with $\bar{n}^\mu$ and $\bar{t}^\mu$ replacing $n^\mu$. In SCET, the gluon fields in the Wilson line are soft gluons for the $Y$’s and collinear gluons for the $W$’s, but once the sectors are decoupled one can treat any of these gluons simply as a gluon field of full QCD.

The hard function is the square of the Wilson coefficient:

$$H(Q^2, \mu) = |C(Q^2, \mu)|^2. \quad (B.2.3)$$

While we do not present its expression here, the hard function for $e^+e^- \rightarrow gg$ events is defined analogously, by matching the Higgs current $F_{\mu\nu}F^{\mu\nu}$ onto SCET.

**B.2.1** $e^+e^- \rightarrow q\bar{q}$

The one-loop hard function for the process $e^+e^- \rightarrow q\bar{q}$ is [53, 57, 170, 264]

$$H = 1 + \frac{\alpha_s C_F}{2\pi} \left(-L_H^2 - 3L_H - 8 + \frac{7}{6} \pi^2 \right), \quad (B.2.4)$$

314
where
\[ L_H = \log \frac{\mu^2}{Q^2}. \quad (B.2.5) \]

The cusp anomalous dimension of the hard function to all orders is

\[ \Gamma_H = -2C_F \Gamma_{\text{cusp}}, \quad (B.2.6) \]

where \( \Gamma_{\text{cusp}} \) is the cusp anomalous dimension defined in Eq. (B.1.5). Similar to the cusp anomalous dimension, we define the coefficients of the non-cusp anomalous dimension \( \gamma \) via

\[ \gamma = \sum_{n=0}^{\infty} \gamma^{(n)} \left( \frac{\alpha_s}{4\pi} \right)^{n+1}. \quad (B.2.7) \]

Through two-loops, the non-cusp anomalous dimension coefficients of the hard function are [266, 307]

\[ \gamma^{(0)}_H = -12C_F, \quad (B.2.8) \]

\[ \gamma^{(1)}_H = \left( -6 + 8\pi^2 - 96\zeta_3 \right) C_F^2 + \left( -\frac{1922}{27} - \frac{22}{3}\pi^2 + 104\zeta_3 \right) C_FC_A \]

\[ + \left( \frac{520}{27} + \frac{8}{3}\pi^2 \right) C_F n_f T_R. \quad (B.2.9) \]
B.2.2 \( e^+e^- \rightarrow gg \)

In the infinite top quark mass limit or with a finite Yukawa coupling, \( e^+e^- \) scattering can produce final state gluon jets. The hard function for such a process can be extracted from \( gg \rightarrow H \) calculations. To all orders, the cusp anomalous dimension of the \( e^+e^- \rightarrow gg \) hard function is

\[
\Gamma_H = -2C_A\Gamma_{\text{cusp}}, \quad (B.2.10)
\]

where \( \Gamma_{\text{cusp}} \) is the cusp anomalous dimension in Eq. (B.1.5). Through two-loops, the coefficients of the non-cusp anomalous dimension are [35, 213, 289]

\[
\gamma_H^{(0)} = -4\beta_0, \quad (B.2.11)
\]

\[
\gamma_H^{(1)} = \left( -\frac{236}{9} + 8\zeta_3 \right) C_A^2 + \left( -\frac{76}{9} + \frac{2}{3}\pi^2 \right) C_A\beta_0 - 4\beta_1.
\]

B.3 The Global Soft Function

For arbitrary exponent \( \beta \) in the soft-drop groomer, the soft function can be calculated by requiring that soft gluons in measured jets fail the soft drop criterion. For hemisphere jets in \( e^+e^- \rightarrow q\bar{q} \) events, for example, the soft function is defined by the
forward matrix element of soft Wilson lines:

\[ S_G(z_{\text{cut}}) = \frac{1}{N_C} \text{tr}(0|T\{Y_nY_{\bar{n}}\}\hat{\Theta}_{SD}T\{Y_nY_{\bar{n}}\}|0). \] (B.3.1)

Here, \( n \) and \( \bar{n} \) are the light-like directions of the \( q\bar{q} \) dipole, \( T \) denotes time ordering, and \( \hat{\Theta}_{SD} \) denotes the soft drop groomer operator which requires the final state to fail soft drop. The action of \( \hat{\Theta}_{SD} \) on soft final states cannot be written in a closed form for an arbitrary final state due to clustering effects, though it can be defined order-by-order. For example, the matrix element of \( \hat{\Theta}_{SD} \) for \( \beta = 0 \) on a final state with two soft particles was presented in Sec. 3.5.1.1.

At one-loop for hemisphere jets in \( e^+e^- \) collisions, the soft function \( S_G \) can be calculated from

\[ S_G = g^2 \mu^{2\varepsilon}C_i \int \frac{d^4k}{(2\pi)^d} \frac{n \cdot \bar{n}}{n \cdot k} \cdot \frac{2\pi\delta(k^2)\Theta(k^0)\Theta(\bar{n} \cdot k - n \cdot k)\Theta(z_{\text{cut}} Q/2 \left[ 2 \frac{n \cdot k}{k^0} \right]^{\beta/2} - k^0)}{\bar{n} \cdot k > \bar{n} \cdot k > 2 n \cdot k \text{ restricts the radiation to lie in one hemisphere, while the requirement}} \] (B.3.2)

where \( n, \bar{n} \) are back-to-back light-like vectors with \( n \cdot \bar{n} = 2 \). The requirement \( \bar{n} \cdot k > n \cdot k \) restricts the radiation to lie in one hemisphere, while the requirement

\[ z_{\text{cut}} Q/2 \left[ 2 \frac{n \cdot k}{k^0} \right]^{\beta/2} > k^0 \] (B.3.3)
restricts the soft gluon to fail soft drop. We find

\[ S_G = 1 + \frac{\alpha_s C_i}{\pi} \left[ \frac{1}{2(1 + \beta)} L_S^2 - \frac{\pi^2}{12} \left( \frac{1}{1 + \beta} + 2 + \beta \right) \right], \quad (B.3.4) \]

where \( C_i \) is the appropriate color factor (\( C_F \) for \( e^+ e^- \to q\bar{q} \); \( C_A \) for \( e^+ e^- \to gg \)) and

\[ L_S = \log \frac{\mu^2}{Q^2(z_{\text{cut}})^2 4\beta}. \quad (B.3.5) \]

To all orders, the cusp anomalous dimension of the hemisphere wide-angle soft function is

\[ \Gamma_S = \frac{2C_i}{1 + \beta} \Gamma_{\text{cusp}}, \quad (B.3.6) \]

where \( \Gamma_{\text{cusp}} \) is the cusp anomalous dimension from Eq. (B.1.5). To one-loop order, the non-cusp anomalous dimension is 0:

\[ \gamma_S^{(0)} = 0. \]

For NNLL resummation, we need the non-cusp anomalous dimension to two-loop order. As discussed in Sec. 3.5.1.1, for soft drop with angular exponent \( \beta = 0 \), this can be extracted from energy veto calculations, up to clustering effects that we calculated. For soft drop with \( \beta = 0 \) and Cambridge/Aachen reclustering, we find the
two-loop non-cusp anomalous dimension to be

\[
\left. \gamma^{(1)}_S \right|_{\beta=0} = C_i \left[ 34.01 \, C_F + \left( \frac{1616}{27} - 56 \zeta_3 - 9.31 \right) \, C_A - \left( \frac{448}{27} + 14.04 \right) \, n_f T_R - \frac{2\pi^2}{3} \beta_0 \right].
\]

(B.3.7)

\section*{B.4 Jet Functions}

Here, we present the quark and gluon jet functions on which the energy correlation function \( e_2^{(\alpha)} \) is measured. The quark jet function, for example, is defined by the forward matrix element:

\[
J_q(e_2^{(\alpha)}) = \frac{(2\pi)^3}{N_C} \text{tr}(0) \, \frac{\bar{\chi}}{2} \chi_{n}(0) \delta(Q - \hat{n} \cdot \mathcal{P}) \delta^{(2)}(\vec{P}_\perp) \delta \left( e_2^{(\alpha)} - e_2^{(\alpha)} \right) \bar{\chi}_n(0) |0\rangle.
\]

(B.4.1)

Here, the jet is collinear to the light-like direction \( n \), the operator \( \delta(Q - \hat{n} \cdot \mathcal{P}) \) restricts the large light-cone component of momentum to be equal to the center-of-mass collision energy \( Q \), and \( \delta^{(2)}(\vec{P}_\perp) \) restricts the jet function to have zero net momentum transverse to the \( n \) direction. The measurement operator is defined by its action on an \( n \)-particle collinear final state \( |X_n\rangle \) as:

\[
\hat{e}^{(\alpha)}_2 |X_n\rangle = \frac{2^{3\alpha/2}}{Q^2} \sum_{i<j \in X_n} (\hat{n} \cdot p_i)^{1-\alpha/2}(\hat{n} \cdot p_j)^{1-\alpha/2}(p_i \cdot p_j)^{\alpha/2} |X_n\rangle.
\]

(B.4.2)
To write this expression, we have expanded the definition of the energy correlation
function from Sec. 3.2.2 to leading power with collinear momenta. The gluon jet func-
tion is defined similarly:

\[
J_g(e_2^{(\alpha)}) = \frac{(2\pi)^3}{N_C} \text{tr}(0|B^\mu_\perp(0)\delta(Q - \vec{n} \cdot \vec{P})\delta^{(2)}(\vec{F}_\perp)\delta\left(e_2^{(\alpha)} - \hat{e}_2^{(\alpha)}\right)B_{\perp \mu}(0)|0\rangle ,
\]

where \( B^\mu_\perp \) is the collinear-gauge invariant operator in SCET that creates physical
collinear gluons.

The following expressions will be presented in Laplace space, where renormalization
is multiplicative and the Laplace space conjugate is \( \nu \). That is,

\[
J(\nu) = \int_0^{\infty} de_2^{(\alpha)} e^{-\nu e_2^{(\alpha)}} J(e_2^{(\alpha)}) .
\]

The one-loop quark and gluon jet functions were first calculated in Ref. [251] for jets
on which the two-point energy correlation functions with arbitrary angular exponent
are measured.
### B.4.1 Quark Jets

To one loop, the Laplace-space quark jet function is

\[
J_q(\nu) = 1 + \frac{\alpha_s C_F}{2\pi} \left[ \frac{\alpha}{2(\alpha - 1)} L_C^2 + \frac{3}{2} L_{C} + \left( \frac{13}{2} - \frac{12}{2\alpha} \right) \frac{\pi^2}{12} \left( 9 - \frac{3}{\alpha - 1} - \frac{4}{\alpha} \right) \right],
\]

(B.4.5)

where

\[
L_C = \log \frac{\mu^2 (\nu e^{\gamma_E})^{2/\alpha}}{E_J^2}.
\]

(B.4.6)

To all orders, the cusp anomalous dimension of the quark jet function is

\[
\Gamma_C^q = \frac{\alpha}{\alpha - 1} C_F \Gamma_{\text{cusp}},
\]

(B.4.7)

where \(\Gamma_{\text{cusp}}\) is the cusp anomalous dimension from Eq. (B.1.5). For all \(\alpha\), the one-loop non-cusp anomalous dimension is

\[
\gamma_{C}^{q,(0)} = 6C_F.
\]

(B.4.8)

For NNLL resummation, we also need the two-loop non-cusp anomalous dimension.

For \(\alpha = 2\), corresponding to jet mass or thrust, this is known exactly. In that case, the non-cusp anomalous dimension is [277]

\[
\gamma_{C}^{q,(1)} \bigg|_{\alpha=2} = C_F \left[ C_F \left( 3 - 4\pi^2 + 48\zeta_3 \right) + C_A \left( \frac{1769}{27} + \frac{22\pi^2}{9} - 80\zeta_3 \right) \right]
\]

321
\[ + T_{Rn_f} \left( -\frac{484}{27} - \frac{8\pi^2}{9} \right) \]. \hspace{1cm} (B.4.9)

**B.4.2 Gluon Jets**

To one-loop, the Laplace-space gluon jet function is

\[ J_g(\nu) = \hspace{1cm} (B.4.10) \]

\[ 1 + \frac{\alpha_s}{2\pi} \left[ \frac{\alpha C_A}{2(\alpha - 1)} L_C^2 + \frac{\beta_0}{2} L_C + C_A \left( \frac{67\alpha - 1}{9\alpha} - \frac{\pi^2}{3} \frac{2(\alpha - 1)^2 - 1}{\alpha - 1} \right) + n_f T_R \left( \frac{26}{9\alpha} - \frac{23}{9} \right) \right], \]

where

\[ L_C = \log \frac{\mu^2 (\nu e^\gamma E_f)^{2/\alpha}}{E_f^2}. \hspace{1cm} (B.4.11) \]

To all orders, the cusp anomalous dimension of the gluon jet function is

\[ \Gamma_C^g = \frac{\alpha}{\alpha - 1} C_A \Gamma_{\text{cusp}}, \hspace{1cm} (B.4.12) \]

where \( \Gamma_{\text{cusp}} \) is the cusp anomalous dimension from Eq. (B.1.5). For all \( \alpha \), the one-loop non-cusp anomalous dimension is

\[ \gamma_C^{g,(0)} = 2\beta_0. \hspace{1cm} (B.4.13) \]

For NNLL resummation, we also need the two-loop non-cusp anomalous dimension.
For $\alpha = 2$, corresponding to jet mass or thrust, this is known exactly. In that case, the non-cusp anomalous dimension is [70]

$$\gamma_C^{(1)} = C_A^2 \left( \frac{2192}{27} - \frac{22\pi^2}{9} - 32\zeta_3 \right) + C_A T_R n_f \left( -\frac{736}{27} + \frac{8\pi^2}{9} \right) - 8C_F T_R n_f. \quad (B.4.14)$$

### B.5 Collinear-Soft Function

The final piece in the factorization theorem is the collinear soft function, defined from soft radiation that is collinear to the jet. As it describes soft radiation, the collinear-soft function is defined as a forward matrix element of Wilson lines:

$$S_C(z_{cut} e_2^{(\alpha)}) = \frac{1}{N_C} \text{tr}\langle 0 | T \{ Y_n^I W_I \} \delta \left( e_2^{(\alpha)} - \left( 1 - \hat{\Theta}_{SD} \right) \hat{e}_2^{(\alpha)} \right) T \{ W_I^I Y_n^I \} | 0 \rangle. \quad (B.5.1)$$

The $Y$ and $W$ Wilson lines are the same as the ones in the soft and jet functions respectively, but depend on collinear-soft fields (which, like any of the others, can be treated as full QCD fields at leading power).

Now, collinear-soft modes only contribute to $e_2^{(\alpha)}$ if emissions pass the soft drop groomer: this is denoted by $1 - \hat{\Theta}_{SD}$ in the measurement function. (Recall that $\hat{\Theta}_{SD}$ removes emissions from the jet according to soft drop.) Again, this operator cannot be written in closed form for an arbitrary final state due to clustering effects, but below, we will calculate it explicitly at one-loop. The $\hat{e}_2^{(\alpha)}$ measurement operator is
defined by its action on an $n$-particle collinear-soft final state $|X_{S,n}|$

$$
\hat{e}_2^{(\alpha)} |X_{S,n}\rangle = \frac{2^{\alpha}}{Q} \sum_{i \in X_{S,n}} (\bar{n} \cdot p_i)^{1-\alpha/2} (n \cdot p_i)^{\alpha/2} |X_{S,n}\rangle. \tag{B.5.2}
$$

This follows from expanding the definition of the energy correlation function from Sec. 3.2.2 to leading power with collinear-soft momenta.

This can be calculated at one-loop accuracy from

$$
S_C = g^2 \mu^{2\epsilon} C_i \int \frac{d^d k}{(2\pi)^d} \frac{n \cdot \bar{n}}{n \cdot k \cdot \bar{n}} 2\pi \delta(k^2) \Theta(\bar{n} \cdot k) \left[ \Theta \left( z_{\text{cut}} \left[ \frac{4 \bar{n} \cdot k}{\bar{n} \cdot \bar{k}} \right]^{\beta/2} - \frac{\bar{n} \cdot k}{Q} \right) \delta \left( e_2^{(\alpha)} \right) 
\right.
\left. + \Theta \left( \frac{\bar{n} \cdot k}{Q} - z_{\text{cut}} \left[ \frac{4 \bar{n} \cdot k}{\bar{n} \cdot \bar{k}} \right]^{\beta/2} \right) \delta \left( e_2^{(\alpha)} - \frac{2^{\alpha}}{Q} (n \cdot k)^{\alpha/2} (\bar{n} \cdot k)^{1-\alpha/2} \right) \right]

= g^2 \mu^{2\epsilon} C_i \int \frac{d^d k}{(2\pi)^d} \frac{n \cdot \bar{n}}{n \cdot k \cdot \bar{n}} 2\pi \delta(k^2) \Theta(\bar{n} \cdot k) \Theta \left( \frac{\bar{n} \cdot k}{Q} - z_{\text{cut}} \left[ \frac{4 \bar{n} \cdot k}{\bar{n} \cdot \bar{k}} \right]^{\beta/2} \right)
\times \left[ \delta \left( e_2^{(\alpha)} - \frac{2^{\alpha}}{Q} (n \cdot k)^{\alpha/2} (\bar{n} \cdot k)^{1-\alpha/2} \right) - \delta \left( e_2^{(\alpha)} \right) \right]. \tag{B.5.3}
$$

where $C_i$ is the color factor of the jet. In the second equality, we have rearranged the phase space constraints and explicitly removed scaleless integrals. For this collinear-soft function, at one-loop in Laplace space we find

$$
S_C(\nu) = 1 + \frac{\alpha_s C_i}{2\pi} \left[ \frac{\alpha + \beta}{2(\alpha - 1)(\beta + 1)} L_{S_C}^2 + \frac{\pi^2}{12} \frac{(\alpha + 2 + 3\beta)(\alpha - 2 - \beta)}{(\alpha + \beta)(\alpha - 1)(\beta + 1)} \right], \tag{B.5.4}
$$

324
where
\[
L_{SC} = \log \frac{\mu^2 (\nu e^{\gamma_E})^{2^{\frac{\beta+1}{\alpha+\beta}}}}{E_z^2 (z_{\text{cut}})^{2^{\frac{\alpha-1}{\alpha+\beta}}}}.
\] (B.5.5)

To all orders, the cusp anomalous dimension of the collinear-soft function is
\[
\Gamma_{SC} = -C_i \frac{\alpha + \beta}{(\alpha - 1)(\beta + 1)} \Gamma_{\text{cusp}},
\] (B.5.6)

where \(\Gamma_{\text{cusp}}\) is the cusp anomalous dimension from Eq. (B.1.5). To one-loop order, the non-cusp anomalous dimension is 0:
\[
\gamma_{SC}^{(0)} = 0.
\]

For NNLL resummation, we need the non-cusp anomalous dimension to two-loop order. For \(\alpha = 2\) and \(\beta = 0\), this can be determined by renormalization group consistency of the cross section directly, using either the \(e^+e^- \rightarrow q\bar{q}\) or the \(e^+e^- \rightarrow gg\) process. For soft drop with Cambridge/Aachen reclustering, the two-loop non-cusp anomalous dimension is
\[
\gamma_{SC}^{(1)} \bigg|_{\alpha=2,\beta=0} = C_i \left[ -17.00 C_F + \left( -55.20 + \frac{22\pi^2}{9} + 56\zeta_3 \right) C_A + \left( 23.61 - \frac{8\pi^2}{9} \right) n_f T_R \right].
\] (B.5.7)
B.6 Resummation

Because we work in Laplace space, defined according to

\[ F(\nu) = \int_{0}^{\infty} d\epsilon_2^{(\alpha)} e^{-\nu \epsilon_2^{(\alpha)}} F(\epsilon_2^{(\alpha)}), \]  

(B.6.1)

the renormalization of all functions in the factorization theorem is multiplicative. For some function \( F \) in the factorization theorem, it generically has the renormalization equation

\[ \mu \frac{\partial}{\partial \mu} F(\mu) = \gamma F(\mu), \]  

(B.6.2)

where the anomalous dimension of \( F \) is \( \gamma \). The anomalous dimension can be written as

\[ \gamma = \Gamma_F(\alpha_s) \log \frac{\mu^2}{\mu_1^2} + \gamma_F(\alpha_s), \]  

(B.6.3)

where \( \Gamma_F(\alpha_s) \) is the cusp part of the anomalous dimension, \( \mu_1 \) is the infrared scale in the logarithm and \( \gamma_F(\alpha_s) \) is the non-cusp part of the anomalous dimension. The solution\(^1\) to the renormalization group equation can be written more conveniently as

\[ \Gamma_F(\alpha_s) = \log \frac{\mu^2}{\mu_1^2} + \gamma_F(\alpha_s), \] 

where \( \mu_0 = 1 \text{ GeV} \), \( \mu_1 = \text{infrared scale} \), and \( \mu_2 = \text{mass scale} \). The constant \( \gamma_F(\alpha_s) \) is the anomalous dimension of \( F \) at \( \mu = \mu_0 \).

\[^1\text{In the plots of resummed distributions in this paper, we have frozen the strong coupling at } \mu_{NP} = 1 \text{ GeV to keep cross sections finite. In the case of frozen } \alpha_s, \text{ the solution to the renormalization group equation for each } F(\mu) \text{ is quite simple, so we omit the details of the prescription below } \mu_{NP} \text{ here.}\]
an integral with respect to $\alpha_s$, by using the definition of the $\beta$-function as

$$\frac{d\mu}{\mu} = \frac{d\alpha_s}{\beta(\alpha_s)}. \quad (B.6.4)$$

Then, the solution to Eq. (B.6.3) can be expressed as

$$F(\mu) = F(\mu_0) \exp \left[ 2 \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)} \Gamma_F(\alpha) \int_{\alpha_s(\mu_0)}^{\alpha} \frac{d\alpha'}{\beta(\alpha')} + \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)} \gamma_F(\alpha) \right. \quad (B.6.5)

$$

$$+ \log \frac{\mu^2}{\mu^2_1} \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)} \Gamma_F(\alpha) \right],$$

where $\mu_0$ is a reference scale.

The exponentiated kernels can be explicitly evaluated to any logarithmic accuracy given the anomalous dimensions. The cusp-part of the anomalous dimension, $\Gamma_F(\alpha_s)$, is proportional to the cusp anomalous dimension, $\Gamma_F(\alpha_s) = d_F \Gamma_{cusp}$, where $d_F$ includes an appropriate color factor. The cusp anomalous dimension has an expansion in $\alpha_s$ given by Eq. (B.1.5). The non-cusp anomalous dimension has a similar expansion defined in Eq. (B.2.7). For resummation to NNLL accuracy, we need the $\gamma_0$ and $\gamma_1$ coefficients, corresponding to computing the anomalous dimensions of the functions in the factorization theorem to two-loops.

With these expansions, we are able to explicitly evaluate the exponentiated kernel
to NNLL accuracy. We have:

\[
K_F(\mu, \mu_0) \equiv 2 \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)} \Gamma_F(\alpha) \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha'}{\beta(\alpha')} + \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)} \gamma_F(\alpha) \tag{B.6.6}
\]

\[
= C_i \frac{\Gamma_0}{2\beta_0^2} \left\{ \frac{4\pi}{\alpha_s(\mu_0)} \left( \log r + \frac{1}{r} - 1 \right) + \left( \frac{\Gamma_1}{\Gamma_0} - \frac{\beta_1}{\beta_0} \right) \left( r - 1 - \log r \right) - \frac{\beta_1}{2\beta_0} \log^2 r 
+ \frac{\alpha_s(\mu_0)}{4\pi} \left[ \left( \frac{\Gamma_1}{\Gamma_0} - \frac{\beta_1}{\beta_0} \right) \left( r - 1 - r \log r \right) - \left( \frac{\beta_1^2}{\beta_0^2} - \frac{\beta_2}{\beta_0} \right) \log r 
+ \left( \frac{\Gamma_2}{\Gamma_0} - \frac{\Gamma_1}{\Gamma_0} \right) + \frac{\beta_2}{\beta_0} \right] \frac{r^2 - 1}{2} + \left( \frac{\Gamma_2}{\Gamma_0} - \frac{\Gamma_1}{\Gamma_0} \right) \Gamma_1 \beta_0 \right\} \right\} 
- \frac{\gamma_0}{2\beta_0} \log r - \frac{\gamma_0}{2\beta_0} \frac{\alpha_s(\mu_0)}{4\pi} \left( \frac{\gamma_1}{\gamma_0} - \frac{\beta_1}{\beta_0} \right) (r - 1),
\]

where

\[
r = \frac{\alpha_s(\mu)}{\alpha_s(\mu_0)}.
\]

The other exponentiated factor is

\[
\omega_F(\mu, \mu_0) \equiv \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{d\alpha}{\beta(\alpha)} \Gamma_F(\alpha) \tag{B.6.7}
\]

\[
= -C_i \frac{\Gamma_0}{2\beta_0} \left\{ \log r + \frac{\alpha_s(\mu_0)}{4\pi} \left( \frac{\Gamma_1}{\Gamma_0} - \frac{\beta_1}{\beta_0} \right) \left( r - 1 \right) 
+ \frac{1}{2} \frac{\alpha_s^2(\mu_0)}{(4\pi)^2} \left( \frac{\beta_1^2}{\beta_0^2} - \frac{\beta_2}{\beta_0} + \frac{\Gamma_2}{\Gamma_0} - \frac{\Gamma_1}{\Gamma_0} \beta_0 \right) \left( r^2 - 1 \right) \right\}.
\]

Then, we can write the solution in Laplace space to the renormalization group equa-
tion in Eq. (B.6.5) as

\[ F(\mu) = e^{K_F(\mu, \mu_0)} F(\mu_0) \left( \frac{\mu_0^2}{\mu_1^2} \right) \omega_F(\mu, \mu_0). \]  

(B.6.8)

Because the hard function and the wide-angle soft function are independent of the observable \( e_2^{(\alpha)} \), their renormalization group equations are identical in real space and Laplace conjugate space. For the jet functions and the collinear-soft function, the inverse Laplace transform is non-trivial.

For any of the jet functions appearing in the factorization theorem, the Laplace space solution can be written as

\[ J(\nu, \mu) = e^{K_J(\nu, \mu_0)} J(\nu, \mu_0) \left[ \frac{\mu_0^2}{E_f} (\nu e^{\gamma_E})^{2/\alpha} \right] \omega_J(\mu, \mu_0). \]  

(B.6.9)

Note that the logarithms that appear in the low-scale jet function \( J(\nu, \mu_0) \) have the same argument as the factor that is raised to the \( \omega_J \) power. Therefore, using the relationship (noted by Ref. [66])

\[ \frac{\partial^n}{\partial q^n} \nu^q = \nu^q \log^n \nu, \]  

(B.6.10)

we can re-write the jet function as

\[ J(\nu, \mu) = e^{K_J(\nu, \mu_0)} J(L \rightarrow \omega_J) \left[ \frac{\mu_0^2}{E_f} (\nu e^{\gamma_E})^{2/\alpha} \right] \omega_J(\mu, \mu_0). \]  

(B.6.11)
Here \( J(L \to \partial \omega_j) \) means that the logarithms in the low-scale jet function \( J(\nu, \mu_0) \) are replaced by derivatives with respect to the exponentiated factor \( \omega_J(\mu, \mu_0) \). The exact same replacement can be made for the collinear-soft function. In that case, we have

\[
S_C(\nu, z_{\text{cut}}, \mu) = e^{K_{SC}(\mu, \mu_0)} S_C(L \to \partial \omega_{SC}) \left[ \frac{\mu_0^2 (\nu e^{\gamma E})^{\frac{2 \beta + 1}{\alpha + \beta}}}{E_J^2 (z_{\text{cut}})^{\frac{2}{\alpha + \beta}}} \right] \omega_{SC}(\mu, \mu_0).
\] (B.6.12)

This re-writing of the jet and collinear-soft functions allows for very straightforward inverse Laplace transformation. In Laplace space, the total differential cross section for left and right hemisphere jets in \( e^+ e^- \) collisions is

\[
\sigma(\nu) = \exp \left[ K_H(\mu, \mu_H) + K_S(\mu, \mu_S) + K_{SC}(\mu, \mu_{SC}^{(L)}) + K_{SC}(\mu, \mu_{SC}^{(R)}) + K_J(\mu, \mu_J^{(L)}) + K_J(\mu, \mu_J^{(R)}) \right] \\
\times H(Q, \mu_H) \cdot S(z_{\text{cut}}, \mu_S) S_C(L \to \partial \omega_{SC}) S_C(L \to \partial \omega_{SC}) J(L \to \partial \omega_J^{(L)}) J(L \to \partial \omega_J^{(R)}) \\
\times \left[ \frac{\mu_H^2}{Q^2} \right] \omega_{H}(\mu, \mu_H) \left[ \frac{\mu_S^2}{4 \frac{z_{\text{cut}}}{Q^2}} \right] \omega_{S}(\mu, \mu_S) \left[ \frac{\mu_{SC}^{(L)}^2 (\nu e^{\gamma E})^{\frac{2 \beta + 1}{\alpha + \beta}}}{E_J^2 (z_{\text{cut}})^{\frac{2}{\alpha + \beta}}} \right] \omega_{SC}(\mu, \mu_{SC}^{(L)}) \\
\times \left[ \frac{\mu_{SC}^{(R)}^2 (\nu e^{\gamma E})^{\frac{2 \beta + 1}{\alpha + \beta}}}{E_J^2 (z_{\text{cut}})^{\frac{2}{\alpha + \beta}}} \right] \omega_{SC}(\mu, \mu_{SC}^{(R)}) \left[ \frac{\mu_J^{(L)}^2 (\nu e^{\gamma E})^{2/\alpha}}{E_J^2} \right] \omega_{J}(\mu, \mu_J^{(L)}) \\
\times \left[ \frac{\mu_J^{(R)}^2 (\nu e^{\gamma E})^{2/\alpha}}{E_J^2} \right] \omega_{J}(\mu, \mu_J^{(R)})
\]

Note that the inverse Laplace transform commutes with the derivatives, and we have

\[
\mathcal{L}^{-1}[\nu^q] = \frac{\Gamma(-q)}{\Gamma(q)}.
\] (B.6.14)
Therefore, the differential cross section in real space can be written as:

\[
e^{(\alpha)} e^{(\alpha)} \frac{d^2 \sigma}{d \ell^2, \ell^2} = \exp \left[ K_H(\mu, \mu_H) + K_S(\mu, \mu_S) + K_{SC}(\mu, \mu_{SC}^{(L)}) + K_{SC}(\mu, \mu_{SC}^{(R)}) + K_J(\mu, \mu_J^{(L)}) + K_J(\mu, \mu_J^{(R)}) \right] \times H(Q, \mu_H) S(z_{cut}, \mu_S) \times \omega_{\omega_{SC}}^{(L)} S_{C}(L \rightarrow \partial_{\omega_{SC}}^{(L)} J(L \rightarrow \partial_{\omega_j}^{(R)}) J(L \rightarrow \partial_{\omega_j}^{(R)}) \times \left( \frac{\mu_{H}^2}{Q^2} \right)^{\omega_{\omega_{H}}(\mu, \mu_H)} \left( \frac{\mu_{S}^2}{4^\beta Z_{\text{cut}}^2 Q^2} \right)^{\omega_{\omega_{S}}(\mu, \mu_S)} \frac{(\mu_{SC}^{(L)})^2 (e_{2, L}^{(\alpha)} e^{-\gamma E})^{-2\beta + 1}}{E_J^2 (z_{cut})^2 \frac{\alpha - 1}{\alpha + \beta}} \left( \frac{\mu_{SC}^{(R)}}{E_J^2 (z_{cut})^2 \frac{\alpha - 1}{\alpha + \beta}} \right)^{\omega_{\omega_{SC}}^{(L)}} \left[ \frac{(\mu_{J}^{(L)})^2 (e_{2, R}^{(\alpha)} e^{-\gamma E})^{-2\beta + 1}}{E_J^2 (z_{cut})^2 \frac{\alpha - 1}{\alpha + \beta}} \right]^{\omega_{\omega_{J}}^{(L)}} \right. \times \left[ \frac{(\mu_{J}^{(R)})^2 (e_{2, R}^{(\alpha)} e^{-\gamma E})^{-2\beta + 1}}{E_J^2 (z_{cut})^2 \frac{\alpha - 1}{\alpha + \beta}} \right]^{\omega_{\omega_{J}}^{(R)}} \left[ \Gamma \left( -\frac{2(\beta + 1)}{\alpha + \beta} \omega_{\omega_{SC}}^{(L)} - \frac{2}{\alpha} \omega_{\omega_{J}}^{(R)} \right) \right]^{-1} \times \left[ \Gamma \left( -\frac{2(\beta + 1)}{\alpha + \beta} \omega_{\omega_{SC}}^{(R)} - \frac{2}{\alpha} \omega_{\omega_{J}}^{(R)} \right) \right]^{-1}
\]

(B.6.15)

\[\text{B.7 Renormalization Group Evolution of } D_k\]

In this appendix we discuss in detail the renormalization group evolution of the jet flavor coefficient \( D_k \) and explain the procedure we used to estimate the scale uncertainty introduced by neglecting higher-order terms.

The cross section for soft-drop groomed jets in \( pp \rightarrow Z + j \) events factorizes in the

331
\[ \text{limit } e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1, \] where

\[ \frac{d\sigma_{\text{resum}}}{de_2^{(\alpha)}} = \sum_{k=q,\bar{q},g} D_k(p_T^{\text{min}}, \eta_{\text{max}}, z_{\text{cut}}, R) S_{C,k}(z_{\text{cut}} e_2^{(\alpha)}) \otimes J_k(e_2^{(\alpha)}). \quad (B.7.1) \]

The fact that \( D_k \) depends on multiple scales prohibits its resummation to all orders. Nevertheless, its renormalization scale dependence is completely determined by renormalization group invariance of the cross section. We can improve our prediction by solving the following renormalization group equation, which holds at leading power in \( z_{\text{cut}} \):

\[ \frac{\partial \log D_k}{\partial \log \mu} = -\frac{\partial \log(J_k \otimes S_{C,k})}{\partial \log \mu} = \Gamma_{D_k}(\alpha_s) \log \left( \frac{\mu^2}{Q^2} \right) + \gamma_{D_k}(\alpha_s, z_{\text{cut}}), \quad (B.7.2) \]

where \( Q = 2p_{T,J} \). The anomalous dimensions \( \Gamma_{D_k} \) and \( \gamma_{D_k} \) are

\[ \Gamma_{D_k} = -\frac{\beta}{1+\beta} C_k \Gamma_{\text{cusp}}, \quad (B.7.3) \]

\[ \gamma_{D_k} = -(\gamma_{J_k} + \gamma_{S_{C,k}}) - \frac{C_k}{1+\beta} \Gamma_{\text{cusp}} \log z_{\text{cut}}^2. \quad (B.7.4) \]

Here, \( C_k \) is the color Casimir for the jet of flavor \( k \). The anomalous dimension has \( \log z_{\text{cut}} \) dependence, which means \( Q \) is not a natural scale of \( D_k \) where all logarithms are minimized.

332
Nevertheless, we can still formally evolve $D_k$ from a scale $\mu_0 \sim Q$ to a renormalization scale $\mu$ common to the jet and collinear-soft function. Solving the renormalization group evolution Eq. (B.7.2), the improved $D_k$ takes the form

$$D_k(\mu, \mu_f) \equiv D_k(\alpha_s, \mu_0, \mu_f) \left( \frac{\mu_0^2}{Q^2} \right)^{\omega_{D_k}(\mu, \mu_0)} e^{K_{D_k}(\mu, \mu_0)}. \quad (B.7.5)$$

Here, $\mu_f$ represents the factorization scale; i.e., the scale at which the parton distribution functions in $D_k$ are defined. The $\omega_{D_k}$ and $K_{D_k}$ functions are defined in App. B.6. To estimate uncertainties from higher-order corrections due to residual scale dependence in $D_k$, we will vary both $\mu_0$ and $\mu_f$ over the values

$$\mu_0 = \left\{ \frac{Q}{2}, Q, 2Q \right\}, \quad (B.7.6)$$

$$\mu_f = \left\{ \frac{Q}{2}, Q, 2Q \right\}. \quad (B.7.7)$$

For evaluating $D_k$ at fixed-order, we keep the full leading and next-to-leading terms as well as singular terms at the next-to-next-to-leading order in the following expansion of the solution to the renormalization group equation, Eq. (B.7.5). Expanding $D_k$ in powers of $\alpha_s$ as

$$D_k(\alpha_s, \mu, \mu_f) = \sum_{n=0} \left( \frac{\alpha_s(\mu)}{4\pi} \right)^n D_k^{(n)}(\mu, \mu_f), \quad (B.7.8)$$
we have the solutions:

\[ D^{(0)}_k = c^{(0)}_{D_k}, \quad (B.7.9) \]

\[ D^{(1)}_k = \Gamma^{(0)}_{D_k} c^{(0)}_{D_k} \log^2 \frac{\mu}{Q} + \left( \gamma^{(0)}_{D_k} c^{(0)}_{D_k} + 2 \beta_0 c^{(0)}_{D_k} \right) \log \frac{\mu}{Q} + c^{(1)}_{D_k}, \quad (B.7.10) \]

\[ D^{(2)}_k = \frac{1}{2} \left( \Gamma^{(0)}_{D_k} \right)^2 c^{(0)}_{D_k} \log^4 \frac{\mu}{\mu_0} + \left( \gamma^{(0)}_{D_k} \Gamma^{(0)}_{D_k} + \frac{8}{3} \beta_0 \Gamma^{(0)}_{D_k} \right) c^{(0)}_{D_k} \log^3 \frac{\mu}{Q} \]

\[ + \left[ \Gamma^{(1)}_{D_k} + \frac{1}{2} \left( \gamma^{(0)}_{D_k} \right)^2 + 3 \beta_0 \gamma^{(0)}_{D_k} \right] c^{(1)}_{D_k} \log^2 \frac{\mu}{Q} \]

\[ + \left[ \left( \gamma^{(1)}_{D_k} + 2 \beta_1 \right) c^{(0)}_{D_k} + \left( \gamma^{(0)}_{D_k} + 4 \beta_0 \right) c^{(1)}_{D_k} \right] \log \frac{\mu}{Q} + c^{(2)}_{D_k}. \quad (B.7.12) \]

The non-singular terms \( c^{(n)}_{D_k} \) are defined such that at \( \mu = Q \),

\[ D^{(n)}_k (Q, \mu_f) = c^{(n)}_{D_k} (Q, \mu_f). \quad (B.7.14) \]

Therefore one can extract the value of \( c^{(0)}_{D_k} (Q, \mu_f) \) and \( c^{(1)}_{D_k} (Q, \mu_f) \) from MCFM and then extrapolate \( D_k \) to arbitrary value of \( \mu_0 \). Given the current level of precision of MCFM, this procedure can be done through the next-to-leading order. At \( \mathcal{O}(\alpha_s^3) \), the \( c^{(2)}_{D_k} \) term cannot be determined without the next-to-next-to-leading \( pp \rightarrow Z + j \) cross section. Note that the size of \( c^{(2)}_{D_k} \) is no greater than \( \mathcal{O}(\alpha_s^3 \log^4 z_{\text{cut}}) \). Thus we can estimate that the size of uncertainty introduced by the unknown higher-loop non-singular term \( c^{(2)}_{D_k} \) is roughly a factor of \( 1 \pm \alpha_s^2 \log^4 z_{\text{cut}} \), which is beyond NNLL accuracy.
Appendix to Chapter 4
C.1 Weighted Soft Drop Multiplicity

At the end of Sec. 4.3, we used LL reasoning to argue that soft drop multiplicity $n_{SD}$ extracts all of the quark/gluon discriminatory information from the $(z_n, \theta_n)$ variables recorded by ISD. In this appendix, we study a variant of $n_{SD}$, the weighted soft drop multiplicity, defined in Eq. (4.2.8) and repeated for convenience:

$$n_{SD}^{(\kappa)} = \sum_n z_n^\kappa. \quad (C.1.1)$$

While quark/gluon performance is not improved by weighting, the purpose of this appendix is to demonstrate that the techniques of this paper are applicable to a variety of observables.

C.1.1 Discrimination Power

For small values of $\kappa$, the weighted soft drop multiplicity is still sensitive to all emissions in the region $A_{emit}$. On the other hand, as $\kappa \to \infty$, only the largest $z_n$ value contributes significantly to the observable. As a result, the weighted multiplicity interpolates between counting and additive behavior, in the limits $\kappa \to 0$ and $\kappa \to \infty$, respectively. The $\kappa$ dependence of the discrimination power, extracted from VINCIA, is shown in Fig. C.1. One can see that the quark/gluon performance decreases monotonically as $\kappa$ increases.
Figure C.1: Quark gluon discrimination power of weighted soft drop multiplicity as a function of $\kappa$, at the benchmark parameters from Eq. (4.2.7). We also show the limit $\kappa \to \infty$, which is equivalent to $\max(z_n)$.

The LL distribution of the weighted soft drop multiplicity is analytically complicated. Indeed, any analytic expression for it must contain a sum of distributions, one for each value of the number $n$ of counted emissions. For example, when $\beta \leq 0$, each emission contributes at least $z_{\kappa_{\text{cut}}}^\kappa$, so at most $n$ emissions can contribute to $n_{\text{SD}}^{(\kappa)}$ if its value is below $n z_{\kappa_{\text{cut}}}^\kappa$. A full analysis along these lines is carried out in App. C.1.2 below.

To qualitatively understand the trend in Fig. C.1, consider the limit in which ISD records many emissions. Strictly speaking, this analysis is not quantitatively applicable in the perturbative regime, where $n \lesssim 10$ emissions are counted. Nor is this reasoning applicable in the collinear-unsafe regime studied in App. C.1.3, where solely perturbative reasoning is insufficient. Nonetheless, the many-emission limit serves to
build intuition.

In the double-logarithmic approximation, where emissions are soft and collinear
and $\alpha_s$ is a fixed coupling, the weighted multiplicity distribution can be found from
summing independent identically distributed numbers. By the central limit theorem,
this converges to a normal distribution in the limit of many recorded emissions. In
this limit, it suffices to compute the mean and variance of $n_{\mathrm{SD}}^{(\kappa)}$ to estimate its discrimi-
nation power. These are determined at lowest order from the average values of $z\kappa$
and $z^{2\kappa}$ in the allowed emission region as

$$
\langle n_{\mathrm{SD}}^{(\kappa)} \rangle_i = \rho_i A_{\text{emit}} \langle z\kappa \rangle, \quad \text{Var} \left( n_{\mathrm{SD}}^{(\kappa)} \right)_i = \rho_i A_{\text{emit}} \langle z^{2\kappa} \rangle, \quad \text{(C.1.2)}
$$

where

$$
\langle z\kappa \rangle = \frac{1}{A_{\text{emit}}} \int_{\theta_{\text{cut}}}^{R_0} \frac{d\theta}{\theta} \int_{z_{\text{cut}}}^{1/2} \frac{dz}{z} z\kappa \Theta \left[ z - z_{\text{cut}} \left( \frac{\theta}{R_0} \right) ^\beta \right]. \quad \text{(C.1.3)}
$$

With a fixed coupling, the mean value of $z\kappa$ for $\beta > 0$ is

$$
A_{\text{emit}} \langle z\kappa \rangle^{\beta>0} = \frac{1}{2^{\kappa-1} \kappa} \log \frac{R_0}{\theta_{\text{cut}}} - \frac{z_{\text{cut}}^{\kappa}}{\beta\kappa^2} \left( 1 - \left( \frac{\theta_{\text{cut}}}{R_0} \right) ^\beta \right). \quad \text{(C.1.4)}
$$
For $\beta < 0$, the mean value is

$$A_{\text{emit}} \langle z^\kappa \rangle_{\beta < 0} = \Theta \left[ \theta_{\text{cut}} - (2z_{\text{cut}})^{1/\beta} R_0 \right] \left( \frac{1}{2^{\kappa \kappa}} \log \frac{R_0}{\theta_{\text{cut}}} - \frac{z_{\text{cut}}^\kappa}{\beta \kappa^2} \left[ 1 - \left( \frac{\theta_{\text{cut}}}{R_0} \right)^{\beta \kappa} \right] \right) \tag{C.1.5}
$$

$$+ \Theta \left[ (2z_{\text{cut}})^{1/\beta} R_0 - \theta_{\text{cut}} \right] \left( \frac{1}{2^{\kappa \beta \kappa}} \log(2z_{\text{cut}}) - \frac{z_{\text{cut}}^\kappa}{\beta \kappa^2} \left[ 1 - (2z_{\text{cut}})^{-\kappa} \right] \right).$$

Because of the $\rho_i$ prefactor in Eq. (C.1.2), we see that the mean and variance once again satisfy Casimir scaling as in Eq. (4.3.11). Moreover, both the variance and mean scale with the counted area $A_{\text{emit}}$, establishing that the weighted soft drop multiplicity is Poisson-like distributed as defined in Sec. 4.3.

The discrimination power is determined by the relative width

$$w_{\text{rel}} \equiv \sqrt{\frac{\text{Var} \langle n^{(\kappa)}_{\text{SD}} \rangle_i}{\langle n^{(\kappa)}_{\text{SD}} \rangle_i}} = \frac{1}{\sqrt{\rho_i A_{\text{emit}}}} \frac{\sqrt{(z^\kappa)}}{\langle z^\kappa \rangle}. \tag{C.1.6}$$

We can get a sense for the behavior of $w_{\text{rel}}$ by considering two extreme limits. For $\kappa \to 0$ and any choice of $\beta$, the mean value $\langle z^\kappa \rangle$ (and hence $w_{\text{rel}}$) approaches a constant, independent of $\kappa$. For $\kappa \to \infty$, the mean value scales with $\kappa$ like

$$A_{\text{emit}} \langle z^\kappa \rangle_{\kappa \to \infty} \sim \frac{1}{2^{\kappa \kappa}}, \tag{C.1.7}$$
with $z_{\text{cut}} < 1/2$, such that the relative width scales as

$$w_{\text{rel}}^{\kappa \to \infty} \sim \sqrt{\kappa}. \quad (C.1.8)$$

Since the relative width increases with increasing $\kappa$, this reasoning predicts that the discrimination power decreases as $\kappa$ increases. This implies the best discrimination power is attained for $\kappa = 0$ (i.e. ordinary soft drop multiplicity) and decreases for higher $\kappa$. Physically, the discrimination power of $n_{\text{SD}}^{(\kappa)}$ comes from sensitivity to multiple emissions, and for higher $\kappa$, sensitivity to softer emissions is decreased. In the extreme limit of $\kappa \to \infty$, the weighted soft drop multiplicity reduces to the energy fraction of the hardest emission, $n_{\text{SD}}^{\kappa \to \infty} = \max(z_n)$.

This qualitatively explains the trend seen in Fig. C.1, i.e. that the discrimination power monotonically decreasing as $\kappa$ increases. In the limit $\kappa \to \infty$, the discrimination power reaches the universal result predicted by Casimir scaling (slightly off due to small nonperturbative corrections), as the observable $\max(z_n)$ is determined by a Sudakov form factor.

### C.1.2 Analytic Calculation

Using evolution equations similar to those employed in Sec. 4.4, we can compute the distribution of IRC-safe weighted soft drop multiplicities. We will demonstrate this here at LL for simplicity; by taking into account flavor changes and energy losses, one
could obtain NLL evolution equations as in Sec. 4.4.2. Since $n_{\text{SD}}^{(\kappa)}$ is a continuous observable, however, significantly more computation time would be required to compute its NLL distribution, in comparison to the discrete unweighted case.

Let $p^i(n_{\text{SD}}, \theta_{\text{cut}}) \, dn_{\text{SD}}$ denote the differential probability that, given a flavor $i$ jet, its weighted soft drop multiplicity is measured to be $n_{\text{SD}}$. Here, we leave the $z_{\text{cut}}$, $\beta$, and $\kappa$ dependence implicit. Though the weighted soft drop multiplicity does not directly count emissions, it is still useful to keep track of the number of contributing emissions, using

$$p^i(n_{\text{SD}}, \theta_{\text{cut}}) = \sum_{n=0}^{\infty} p^i_n(n_{\text{SD}}, \theta_{\text{cut}}), \quad (C.1.9)$$

where $n$ labels the number of counted emissions as before. If we change the resolution angle from $\theta_{\text{cut}}$ to $\theta_{\text{cut}} - \delta\theta_{\text{cut}}$, then

$$p^i_n(n_{\text{SD}}, \theta_{\text{cut}} - \delta\theta_{\text{cut}}) = p^i_n(n_{\text{SD}}, \theta_{\text{cut}}) \left( 1 - \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} \int_0^{1/2} dz \frac{\alpha_s(z \theta \, p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{\text{SD}}(z, \theta) \right)$$

$$+ \frac{\delta\theta_{\text{cut}}}{\theta_{\text{cut}}} \int_0^{1/2} dz \frac{\alpha_s(z \theta \, p_T)}{\pi} P_{i \rightarrow i}(z) \Theta_{\text{SD}}(z, \theta) p^i_{n-1}(n_{\text{SD}} - z^{\kappa}, \theta_{\text{cut}}). \quad (C.1.10)$$

This leads to a linear differential equation. Instead of the Poisson distribution found in Sec. 4.4.1, the solution in this case is differential in $n_{\text{SD}} = \sum_i z_i^{\kappa}$:

$$p^i_n(n_{\text{SD}}, \theta_{\text{cut}}) = \ldots = \ldots \quad (C.1.11)$$
Figure C.2: LL calculation of weighted soft drop multiplicity distributions with \( \kappa = 1 \), compared to \textsc{Vincia}. The plots have two different sets of ISD parameters which were chosen to display the sharp features characteristic of \( n_{SD}^{(\kappa)} \) in the perturbative regime. The curves shown are the probability distribution functions of \( \log n_{SD}^{(1,0)} \), so that they integrate to one in logarithmic space. The leftmost bin is an underflow bin, showing the probability that no emissions were counted by ISD, such that \( n_{SD}^{(1,0)} = 0 \).

\[
e^{-I_{i \to (\theta_{\text{cut}}, R_0)}} \frac{1}{n!} \left( \prod_{i=1}^{n} \int_{\theta_{\text{cut}}}^{R_0} d\theta_i \int_{0}^{1/2} dz_i \frac{\alpha_s(z_i, \theta_i, p_T)}{\pi} P_{i \to (z_i, \theta_i)} \Theta_{SD}(z_i, \theta_i) \right) \delta \left( n_{SD} - \sum_{i=1}^{n} z_i^\kappa \right)
\]

In the perturbative regime, the behavior of \( n_{SD}^{(\kappa)} \) is most clearly seen on a logarithmic scale. Two example LL distributions are displayed in Fig. C.2 and compared to results from \textsc{Vincia}. In these examples, soft drop parameters were chosen to demonstrate that the sharp features of the \( n_{SD}^{(\kappa)} \) distributions are indeed captured by the LL evolution equations. These sharp features result from the edges of the \( p_n^i (n_{SD}, \theta_{\text{cut}}) \) distributions for different values of \( n \). For example, with \( \beta \leq 0 \), the \( p_n^i (n_{SD}, \theta_{\text{cut}}) \) dis-
Figure C.3: RG evolution of collinear-unsafe weighted soft drop multiplicity with $z_{\text{cut}} = 0.01$ and $\kappa = 1$, for the (a) quark-singlet and (b) gluon cases.

C.1.3 Collinear-Unsafe Evolution

In the case of a collinear-unsafe weighted soft drop multiplicity with $\beta = 0$ and $\theta_{\text{cut}} = 0$, we can apply the methods of Sec. 4.5. Specifically, after extracting the GFF at some RG scale $\mu$, we can use Eq. (4.5.7) with the particular choice $f(z) = z^\kappa$ to predict the upwards evolution. In Fig. C.3, we compare the result of the RG evolution for $z_{\text{cut}} = 0.01$ and $\kappa = 1$ to VINCIA, finding overall good agreement. By eye, one can see that these $\kappa = 1$ distributions do not yield as good separation power as the $\kappa = 0$ distributions shown in Fig. 4.16, though the degree of RG evolution is similar for both the weighted and unweighted cases.
References


349


350


[106] Caola, F., Melnikov, K., Röntsch, R., & Tancredi, L. (2015). QCD corrections to $ZZ$ production in gluon fusion at the LHC.


[121] Chatrchyan, S. et al. (2012a). Measurement of the Production Cross Section for Pairs of Isolated Photons in $pp$ collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 01, 133.


[123] Chatrchyan, S. et al. (2012c). Search for a Higgs boson in the decay channel $H$ to $ZZ(\ast)$ to $q\bar{q}\ell^-\ell^+$ in $pp$ collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1204, 036.


356


360


364


365

[228] Khachatryan, V. et al. (2015a). Comparison of the $Z/\gamma^*+\text{jets}$ to $\gamma+\text{jets}$ cross sections in pp collisions at $\sqrt{s} = 8$ TeV.


366


368


