Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg. Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Peter Schichtel

born in: Heidelberg

Oral examination: October, the 15th, 2014

More Jets in more LHC Searches

Referees: Tilman Plehn

Hans-Christian Schultz-Coulon

Abstract

Multi jet observables are a powerful tool to new physics as well as a boost to standard analysis strategies. We show their use in a reasonably model independent dark matter search and a jet veto Higgs analysis. We find however that, these observables are plagued by huge theoretical uncertainties connected to unphysical scale parameters. In the democratic limit we compute analytically the all order resummed jet spectrum at leading log. It obeys so called staircase scaling. With the help of state of the art Monte Carlo tools we study the jet spectrum features in great detail. In addition we also study so called Poisson scaling. This allows us to develop a data driven strategy to fix the standard model multi jet backgrounds.

Zusammenfassung

Multijetobservablen sind nützliche Werkzeuge in new physics Suchen. Des Weiteren führt ihr Verständnis zu einer Verbesserung von Standardanalysestrategien. Wir zeigen das in einer einigermaßen Model unabhängigen SUSY Suche und einer Jetveto Higgsanalyse. Wir stellen jedoch fest, dass diese Observablen mit großen theoretischen Unsicherheiten belastet sind, die mit unphysikalischen Skalenparameteren verknüpft sind. In einem bestimmten Limes berechnen wir die resummierte Form des Jetspektrums zu führender logarithmischer Ordnung. Wir finden, dass das Spektrum so genanntem staircase scaling folgt. Unter Zuhilfenahme modernster Monte Carlo Simulationen untersuchen wir das Jetspektrum ausführlich. Des weiteren studieren wir auch so genanntes Poisson scaling. Dies Alles erlaubt uns eine datengestützte Strategie für die Standardmodel Multijetuntergründe zu entwickeln.

More Jets in more LHC Searches

Peter Schichtel

Institut für theoretische Physik
Universität Heidelberg

PhD thesis

to Maria

Contents

Why Jets?

Finally – the Higgs! These are exciting times. The nature of electroweak symmetry breaking is about to unfold in front of us [1–4]. It is the last step of a longstanding endeavor we just like to call the standard model and a milestone in the history of science. Of course, there are still plenty of open questions. Will this really be the standard model Higgs [6–10]? How well can we determine all its couplings or better: are we able to reconstruct the Lagrangian from the available data? This is all about Yukawa, electroweak gauge, and self interactions, about the nature of the weak sector. More and more people start to think about the features of future Higgs factories [11] to dwell on the physical properties of this new particle. So why bother with jets?

Despite the fact that certain properties of the Higgs might only be measurable at future linear colliders, the LHC physics program has by no means reached its full potential yet. The machine and the detectors are currently being upgraded [12–16] to deliver and deal with tremendous amounts of energy and data. Not only will this help us to further narrow down the properties of the Higgs [17], but maybe also answer the second big question the LHC was constructed for. Even if the Higgs appears to be standard model like, there are good reasons to search for new physics at the LHC [6][1]. The LHC, however, has always been, and will be even more at higher energies, a jet factory. It is therefor crucial to understand QCD jet production in all its glorious details, if we want to identify and interpret the different sources of the physics we observe.

At the beginning of the LHC era we were all eager for more data. We wanted those statistical uncertainties to shrink. The big experimental collaborations published a great variety of physics analysis in many different search channels [19,20]. With the discovery of the Higgs our data analysis mode starts to change. The LHC data has reached an overwhelming size and some studies start to feel theory limitations. We are in the middle of an era of precision physics at the LHC[2]. QCD computations, especially jet production, however, suffer from huge theoretical uncertainties. We have to control these uncertainties, if we want to use jets as a tool to understand hadron collider physics.

In this thesis we study jet radiation patterns in great detail. It is our goal to show that we can understand jet physics from fundamental principles and put this knowledge to use. This means to really use the information contained in the additional jet radiation, which accompanies every process at the LHC. Therefor, we have the following outline. In chapter we introduce the equations which describe jet evolution. We develop an intuitive ansatz of the physical picture at colliders, which will be the frame of our considerations. Of course, we have to limit the depth of these arguments in a thesis like this. There are many very good introductions into particle theory and collider physics. We give some references for those who like to deepen their knowledge or like to study more mathematical and theoretical aspects. Everything about QFT and the standard model can be found in [5, 21–23]. These books also contain some collider applications. Detailed collider phenomenology introductions are, for example, [24, 25]. We then show in a first example, how the understanding of multi jet rates helps to interpret data in terms of broad new physics classes. In addition we explore the theoretical uncertainties, which plague our computations. To improve our theoretical understanding of jet physics we study jet evolution and construct analytic solutions. We discuss the validity and limitations of these solutions. In chapter 1.3 we show, how this knowledge can be used for a data driven take on jet rates. We then study the application of Fox–Wolfram–Moments built from jets to Higgs analysis. Including the knowledge about jets greatly improves the sensitivity of such studies. In chapter 3.0.7 we recapitulate and conclude.

A few remarks about the research presented in this thesis. I already started studying jets at the LHC in my diploma thesis "Jets in LHC searches" [26][3]. In this time I had the pleasure to work together with Christoph Englert, Steffen Schumann, and Tilman Plehn [27, 28]. The findings and plots from this time will also find their way into this thesis. They will be marked correspondingly,

[1] Even at the moment we see the one and other deviation from the standard model expectation [18]. Rumors point to possible SUSY explanations. Nothing with statistical significance yet. There might be exciting times ahead!

[2] Considering that the LHC is a hadron collider this is an outstanding fact on its own.

[3] Thus motivating the title of this thesis.

of course. The baseline of chapter will be based on [29,30], where Erik Gerwick joined our efforts, and [27]. Text or plots taken from this work are marked beforehand. Note that [29] also contains some collider applications, which we do not present here. The study of jet radiation patterns in WBF Higgs searches was originally initiated by Erik Gerwick, Steffen Schumann, and Tilman Plehn in [31]. The potential of Fox–Wolfram–Moments was independently studied by Catherine Bernaciak, Malte Buschmann, Anja Butter, and Tilman Plehn in [32]. I then had the possibility to join for [33]. Together with [28] this will be the foundation of chapter 1.3. Again text and plots used from this research will be marked correspondingly. I am very grateful to my collaborators from whom I learned a lot. I hope in this way I can present our research in a coherent fashion.

Jet Radiation at the LHC

1.1 Physics at hadron colliders

1.1.1 Scales and colliders

At the LHC it is our goal to put the fundamental symmetries, which, to our knowledge, guide natures interactions, to the test. Symmetries and their spontaneous breaking are at the heart of the standard model and the interactions it describes. The principle idea behind collider physics in general is that of scales in nature. These scales are directly connected to the boundaries of validity of our theoretical frame works. There are fundamentally three different scales in physics[4]: time T, spatial resolution R, and available energy E. However, these scales are all linked to each other by special relativity and quantum mechanics via $T = R = 1/E$. Each development in the history of physics was connected to the push of at least one of these scales beyond its former boundaries[5].

Colliders are a natural choice, because of the enormous amount of energy, which can be focused at a single point. The design energy at the LHC, for example, is 14 TeV. This corresponds to time scales of the order of 10^{-26} seconds. The way nature works at these short times is reflected by the correlations and production rates of the particles hitting the detector, which is built around the interaction point. At the LHC this is done by the two big multi purpose experiments ATLAS [34] and CMS [35]. However, interpreting these patterns is far from trivial. We have to overcome several obstacles before we can interpret the LHC data and think about new physics ideas. The first lies within the detector itself. The primary signals are energy deposits as a function of time and space. These need to be turned and interpreted into physical particles and their four momenta or tracks in space[6]. This highly non trivial procedure is within the realm of experimental particle physics and way beyond the scope of this thesis or the expertise of its author. We rely on our experimental colleagues, which do a fabulous job understanding their detectors, to provide data, which can be interpreted in terms of particles. The particle species which can be in principle directly reconstructed are: Photons, Electrons, Muons, neutral and charged hadrons, if their lifetime is long enough. However, even if we trust the identity and four-momenta[7] of the reconstructed particles we are still far away from understanding what happened at the interaction point. In the language of scales the detector is the outer most part in a chain of processes which connect the short (time) scale physics (also called hard interaction) with the observer. With typical sizes of several meters, we talk about roughly 10^{-8} seconds until the produced particles leave the detector[8] or are absorbed by it. As a matter of fact, however, most known particles are unstable with typical life times shorter then 10^{-10} seconds [36]. Their decays can be well understood from other experiments[9]. This knowledge is collected in decay tables, giving us the decay rates and correlations to the daughter particles. However, these are not the fundamental final and initial states which are familiar to us from quantum field theory [21, 22][10]. Closing in further on the hard interaction, around 10^{-22} seconds, we encounter a dramatic effect: confinement [25]. Being the result of the non abelian nature of the strong interaction, we have to deal with the observation, that there are no free color charges. Quarks and gluons, the fundamental fields describing the strong interaction, also called QCD, are no observable degrees of freedom in nature. Only color neutral states can be observed as final states or be prepared as initial states in any experiment. It is precisely behind this wall built

[4] One could also quote density in this context. For the sake of simplicity we will not think about it here.

[5] Note in this context also, that time T is connected to the origin and history of our universe.

[6] It is worthwhile to note, that there are also other reconstructed objects than particles, for example the energy deposit in the calorimeter only, which can be used to study certain aspects of collider phenomenology.

[7] This is all we can (and need to) know about a particle. Note that spins and polarizations of particles cannot be detected directly at the LHC, but need to be reconstructed from other observables.

[8] It is an astonishing matter of fact the the interaction rate at the LHC lies roughly at the same order of magnitude. This shows the technological maturity of the LHC experiments ATLAS and CMS.

[9] The LHCb experiment [37], for example, studies the decay and properties of B mesons.

[10] This is meant from the perspective of the standard model. Of course, we can describe hadron interactions with effective field theories [22].

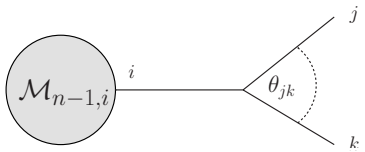

Figure 1.1: Matrix element with a soft respectively collinear structure. We have a total of $n+1$ external lines. Any two of these, say j and k, can always be split apart as shown here.

by confinement, where the physics lives we are interested in. Confinement is a non perturbative effect and cannot be dealt with by our standard text book approaches. However, for physics at shorter time scales than this so called hadronization scale, we know that we can do perturbative QCD and use quarks and gluons as fundamental objects in our computations[11]. This is guaranteed by another astonishing effect of QCD: asymptotic freedom [25]. For shorter and shorter time scales the coupling α_s in QCD approaches zero. Therefor, we are allowed to use perturbative QCD for scales shorter then the hadronization time. Hadronization itself needs to be described by models, which need data input [25]. Lattice computations, which are our theoretical tool to deal with non perturbative QCD, do not have the power to turn a quark gluon distribution into a hadron distribution, yet. However, as will turn out in the following section, jets allow us to ignore most of the effects of hadronization in our considerations. Leaving the realm of very soft[12] physics even shorter time scales can be handled with our standard QFT approaches [5, 21–25].

It is these QFT's, which we want to study and test at the LHC. The theoretical handling of correlation functions in QFT is well understood [21, 22]. Usually we use Feynman graphs to depict the physical process we are interested in. In addition, these graphs serve as computational tool to compute production rates and correlation functions. The standard model is our prime QFT candidate to describe all short scale physics. From other observations, especially concerning the structure of our universe and the matter of fact that gravity is not included in the standard model, as well as theoretical considerations, like the divergent structure in the Higgs sector, we know that it cannot be a complete theory. However, the standard model has not failed us at a particle collider, yet.

1.1.2 What are jets?

We already mentioned that jets play an important role to avoid non perturbative effects. In fact, this whole thesis is about jets. So, what are jets?

There is still some space between the hadronization scale and the scales which we probe at the LHC. Take Higgs production as an example. To produce a Higgs we need at least 126 GeV of energy. Hadronization occurs at approximately 1 GeV. Imagine further, that together with the Higgs we produce a QCD parton. What happens to this parton, when we evolve in time? To understand this we have to take a closer look at the structure of QCD. The fact that QCD (as well as QED) contains mass-less gauge bosons causes a tremendous effect: soft and collinear divergences [22, 24, 25]. Soft divergences are quite intuitive. The amount of energy we have to pay to radiate softer and softer gauge bosons tends to zero. Thus, naively, we can emit infinite many of them. They are an inevitable fact of gauge theories. However, collinear divergences are not that naively accessable. We can understand this fact, if we study the technical situation in which divergences arise. Consider a situation as depicted in Fig. 1.1. The matrix element is

$$\mathcal{M}_{n+1} = \mathcal{M}_{n-1,i}\,\frac{f(p_i)}{p_i^2 - m_i^2}\,V_{ijk}\,\psi_j(p_j)\,\psi_k(p_k). \tag{1.1}$$

With $\psi(p)$ we indicate the external wave functions, where p is the particles momentum. V is the corresponding vertex and f some function of the internal momentum, which might be a unit matrix or just a $\not{p}$ depending on the particle. Let us investigate the propagator in more detail. We use

[11] External states in QFT are defined at infinite time. The fact that the different time scales at which physical processes happen at a collider are factorized by orders of magnitude in time allows us to keep this picture.

[12] We are still orders of magnitude below the available *hard* energy scale at the LHC.

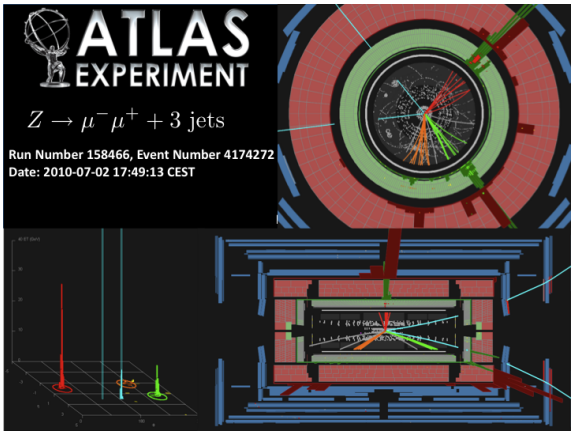

Figure 1.2: Typical event as seen by the ATLAS detector. The two cyan tracks are identified as muons. The red, green, and orange tracks are identified as hadrons. We clearly see their collimated structure. Picture taken from [34].

the implicit δ-function contained in the vertex and find

$$\begin{aligned} p_i^2 - m_i^2 &= m_j^2 + m_k^2 - m_i^2 - 2p_j p_k \\ &\approx m_j^2 + m_k^2 - m_i^2 - 2E_j E_k \left(1 - \left(1 - \frac{m_j^2}{2E_j^2} - \frac{m_k^2}{2E_k^2}\right) \cos\theta_{jk}\right) \\ &\overset{m\to 0}{=} -2E_j E_k \left(1 - \cos\theta_{jk}\right). \end{aligned} \tag{1.2}$$

The propagator diverges for $E_{j,k} \to 0$ (soft) and for $\theta_{jk} \to 0$ (collinear). However, if one of the involved particles is massive the divergence is shielded. In this case there is a center of mass system where the other particles are back to back. Note in addition that only bosons can have a soft divergence due to the Fermi–Dirac statistics. If, however, all three involved particles were gluons, we face a fully divergent structure. In reality we cannot hit these divergent structures as hadronization and detector resolution shield them. In addition, we know that divergent structures point us to inconsistencies in our approach: observable quantities in nature are finite [21, 22]. Nevertheless, this information is enough to know how our parton will evolve. It will radiate other soft and collinear partons, because the divergent structure tells us that exactly this pattern has to be enhanced. These partons then will be the seeds turned into hadrons during hadronization. However, hadronization, living at 1 GeV, has not enough energy available to change the four momenta of the partons too much. This is just Heisenberg's principle: large Δx causes[13] a small Δp. Thus the produced hadrons follow closely the distribution of the radiated partons. This is known as the concept of parton hadron duality [25]. However, their distribution follows from the divergent structure of QCD as argued before. We therefor expect a *collimated spray of hadrons*: this is what we call a jet.

In Fig. 1.2 we show a typical ATLAS event. The hadronic activity is clearly collimated in distinct detector regions. In contrast, in Fig. 1.3 we show a CMS event with huge hadronic activity. The whole detector is flooded with hadrons. In this case we cannot naively associate jets. What we need is a formal procedure to merge hadrons unambiguously into a jet. This is done by so called jet algorithms [24, 25, 38]. A jet algorithm uses a QCD inspired metric and a truncation parameter to define jets. At the LHC we use generalized k_T algorithms with the truncation parameters R and p_T^{2k}. We define the distance between two particles i and j to be

$$d_{ij} = \min\left(p_{T,i}^{2k}, p_{T,i}^{2k}\right) \frac{\Delta R_{ij}}{R}, \tag{1.3}$$

where ΔR_{ij} is the distance in (η, ϕ) space[14]. k defines the specific jet algorithm we use: $k = 1$ corresponds to the k_T algorithm, $k = 0$ is the C/A algorithm, and $k = -1$ corresponds to the anti-k_T algorithm. A jet algorithm than works as follows:

[13] Compared to the hard process our system has expanded over a much larger range in space now.
[14] The parameter R is connected to the geometrical size of a jet.

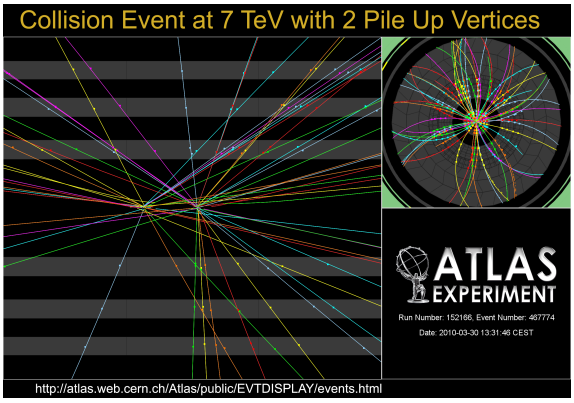

Figure 1.3: An ATLAS event with lots of hadronic activity. The colored tracks are identified hadrons. Picture taken from [34].

1. find $d^{\min} = \min_{ij \in \text{particles}} \left(d_{ij}, p_{T,i}^{2k}\right)$
2. if $d^{\min}$ is of the type $p_{T,i}^{2k}$: call i a jet and remove it from the list of particles, go back to (1.)
3. if $d^{\min}$ is of the type d_{ij}: replace i and j by $i+j$, go back to (1.)
4. iterate until all particles are called jets.

There is one last caveat we have to take care of. Following the previous procedure, we could get arbitrarily soft jets. In addition, arbitrarily collinear jets with respect to the incoming beams would still be allowed, too. However, this would be exactly the region where we would need to take care of the soft and collinear divergence. In addition, it is also the region where hadronization plays a role. An effect we explicitly want to avoid. Therefor, we only keep jets which lie above a certain momentum threshold $p_T^{\min}$, which is far away from the hadronization scale. This way we make sure that the jets we define are really connected to the partons produced in the hard interaction. Typical values in LHC analysis are $R = 0.4 - 0.7$ and $p_T^{\min} = 20 - 100$ GeV.

Jets are our best connection to the partons which built the QCD Lagrangian. What we call a jet in the end depends on the definition of the used jet algorithm. However, once this is fixed, it is unambiguous what belongs to a certain jet. Therefor, studying jet physics is deeply connected to the study of QCD. We will see, that this statement is true, when we take a look at the equations which guide jet evolution in the following sections. This is connected to the fact, that even starting with a single parton we can observe several jets, when some of the branchings happen to be outside the area defined by the jet algorithm.

1.1.3 The factorization theorem

We already argued that the divergent structure of QCD[15] guides the evolution of partons. This structure can be formalized and is known as factorization theorem [22,24,25,39,40]. If two external particles j, k in a process become collinear[16], i.e their separation angle $\theta \to 0$, we can write the cross section as[17]

$$d\sigma_{n+1} = \sum_i d\sigma_{n-1,i} \times \frac{dp_T^2}{p_T^2} dz \frac{\alpha_s}{2\pi} P_{i \to jk}(z), \tag{1.4}$$

where z is the energy fraction of j with respect to i and p_T, the transverse momentum of j, is defined relative to the direction of i and is therefor connected to the collinear angle θ. The total number of

[15] The same arguments are true for QED. However, due to the fact, that the photon carries no charge and is therefor not self interacting, the phenomenology is much simpler [22].

[16] Some care has to be taken if one of the particles is in the initial state. The line of arguments, however, is not touched.

[17] There is also a version starting with the soft limit $E_i \to 0$. However, the collinear version is our choice. More advanced techniques like dipole showers use more involved kinematic considerations to encode both cases as limiting case.

final state particles is $n+1$. Note that the phase space factorizes, too. Therefor this relation holds also for the full cross section [24]. The energy distribution is described by the splitting kernels $P(z)$, which are essentially squared vertices. We compute them in detail in App. A. Dependent on the particular branching the splitting kernels can have divergences for $z \to 0$ or $z \to 1$. These are exactly the soft divergences mentioned earlier. The collinear divergence shows itself in the factor of $1/p_T^2$. This theorem is only exact in the limiting case $\theta \to 0$ or equivalently $p_T \to 0$. For finite p_T, especially for large values, we expect deviations from this behavior[18]. However, integrating over phase space it is clear that the largest part of the cross section will come from regions close to the divergent structure. In addition, for events with a lot of jets we see by purely geometrical considerations, that we have a situation where most jets are collinear to each other.

From the factorized structure we can conclude that we can describe a single radiation with some probabilistic factor. In addition it tells us that we can use this prescription iteratively. Therefor, we can actually describe any number of final state parton configuration. To do this properly we need to find formal evolution equations which give us the radiation pattern generated by a parton when evolving through time. An important observation in this regard is the qualitative behavior of the total cross section when we integrate Eq. (1.4). To avoid the divergence we integrate from some $p_T^{\min}$ to $p_T^{\max}$ [24]. This is exactly the $p_T^{\min}$ used in the definition of the jet algorithm. Our cross section then will be proportional to $\alpha_s \times \log^2(p_T^{\max}/p_T^{\min})$. We get a squared logarithm here because technically p_T is connected to the boundary condition of the integral of the soft divergence, too [25]. There will be also terms with less powers associated to the logarithm. For example coming from the non divergent parts of the splitting kernel. Dependent on the ratio of the phase space boundaries this factor can easily be $\mathcal{O}(1)$ or larger. However, this would mean that our perturbative expansion did not converge at all. We talk about large logarithms occurring in this context. Due to the structure of the factorization theorem it is clear, that adding higher and higher terms does not save us. This is schematically depicted in Tab. 1.1, which shows systematically the logarithms which appear in, and spoil, our computation. A column in this table corresponds to a

	LO	NLO	NNLO	...
LL/DLA	$\alpha_s \log^2$	$\alpha_s^2 \log^4$	$\alpha_s^3 \log^6$	
NDLL		$\alpha_s^2 \log^3$	$\alpha_s^3 \log^5$	
NLL			$\alpha_s^3 \log^4$	
$\vdots$				$\ddots$

Table 1.1: Schematic arrangement of the perturbative series in terms of large phase space logarithms. A standard fixed order computation considers everything in the corresponding column of this table.

fixed order computation, where different logarithms may appear. Each row collects terms with the same quotients between the power of α_s and the logarithm. The technique to solve our problem is known as resummation. We need to sum the whole row depicted with leading log (LL) respectively double logarithmic approximation (DLA) to collect the main effects. Effectively we rearrange the terms of the expansion into a new perturbative series: a logarithmic expansion [24,25]. Terms with a different relation between the logarithm and the power of α_s are sub leading in this expansion and therefor called next to (double) leading logarithms (N(D)LL).

The factorization theorem and the logarithmic considerations following from it are at the heart of all jet radiation considerations. It really is the underlying principle of all following discussions and computations. It is therefor worthwhile to mention that it has been proven only for inclusive DIS and Drell-Yan processes. For all other processes we use it as a working ansatz [41][19].

1.1.4 Inclusive cross section computations

Before we study the details of parton evolution, let us first discuss how actual rates are computed at a hadron collider. The equations and physics which we need to perform this computation are

[18] As a matter of fact we have to deal with a semi classical result, where interference effects are not taken into account. In addition, naively, the distribution in the polar angle is flat. Gluon polarizations are averaged out. However, modern Monte Carlo programs can take care of some of these problems.

[19] This "ansatz" however works astonishingly well.

well known [22, 24, 25]. As a matter of fact, they almost encode everything we need to understand jet evolution and radiation in a much broader context. This will only be clear after detailed discussion though. It is well established, that at hadron colliders, colliding hadrons h_1 and h_2 with a center of mass energy of $\sqrt{s}$, cross sections σ for the final state P are computed via the master equation [22, 24, 25]

$$\sigma(h_1h_2 \to P + X, \sqrt{s}) = \sum_{i,j=q,\bar{q},g} \int dx_1 dx_2 f_{i/h_1}(x_1, \mu_F) f_{j/h_2}(x_2, \mu_F) \times \hat{\sigma}(ij \to P, \mu_F, \mu_R, \sqrt{x_1 x_2 s}). \tag{1.5}$$

This formula encapsulates a lot of physics. Let's begin with its most familiar part. The partonic cross section $\hat{\sigma}$ is computed via Feynman rules at a fixed order[20] in the strong coupling α_s and the electro-weak coupling α. It encodes the probability amplitude of the physical process we are interested in [22]. This is the physics we are after at the LHC. These are the distributions which tell us if we can falsify a model, which we hope to be the standard model in the end. Due to the handling of ultra violet divergences in loop calculations, perturbative cross sections depend on the renormalization scale μ_R, which occurs even at leading order as the argument of the running coupling. These concepts are well known from standard QFT text books [21, 22, 24]. It is the fact, that the hadron in the initial state is not a fundamental particle, where we first encounter a specialty at hadron colliders. This fact is encoded in the parton density function (pdf) $f_{i/h}(x, \mu_F)$. Here f is the probability to find a parton i in the hadron h carrying the momentum fraction x of the hadrons total momentum[21] at a given resolution scale μ_F. From the previous discussion about scales in physics and the divergent structure of QCD it is immediately clear that the apparent field content of the proton will change dramatically if we probe smaller structures. This change is reflected by the famous DGLAP equation[22], which the pdfs have to obey [22, 24, 25]:

$$\begin{aligned} \mu^2 \frac{d}{d\mu^2} f_{i/h}(x, \mu) &= \sum_{i \to jk} \int_x^1 \frac{dz}{z} \frac{\alpha_s}{2\pi} P^+_{i \to jk}(z) f_{j/h}\left(\frac{x}{z}, \mu\right), \\ f_{i/h}(x, \mu) &= \Delta_i(\mu, \mu_0) f_{i/h}(x, \mu_0) + \sum_{i \to jk} \int_{\mu_0}^{\mu} \frac{d\mu'^2}{\mu'^2} \Delta_i(\mu, \mu') \int \frac{dz}{z} \frac{\alpha_s}{2\pi} P_{i \to jk}(z) f_{j/h}\left(\frac{x}{z}, \mu'\right), \end{aligned} \tag{1.6}$$

presented here in its differential and integrated form[23]. $P^{(+)}(z)$ are the (regularized) splitting kernels, mentioned earlier and computed in App. A. The Sudakov form factor Δ_i is given by [22, 24, 25]

$$\Delta_i(\mu, \mu_0) = \exp\left(- \sum_{i \to jk} \int_{\mu_0}^{\mu} \frac{d\mu'^2}{\mu'^2} \int dz \frac{\alpha_s}{2\pi} P_{i \to jk}(z) \right). \tag{1.7}$$

Note that in Eq. (1.6) and (1.7) $i, j, k \in [q, \bar{q}, g]$. It is exactly the second form in Eq. (1.6), which is the key to jet evolution. Technically, the DGLAP equation resums the collinear radiation of the incoming parton. It is therefor a result of the proper treatment of the divergent structure of QCD. The pdfs are non perturbative objects which have to be measured at some reference scale μ_0. The DGLAP equation then tells us how the pdfs change, if we change the energy scale we probe the proton with. The first term in Eq. (1.6) has a simple interpretation: nothing changed. This means no radiation occurred. The Sudakov form factor $\Delta_i(\mu, \mu_0)$ can therefor be interpreted as a non

[20] These computations are non trivial on their own and teach us a lot about QFT. Note that actually the expansion in α is separated into an expansion in the weak coupling and the electro-magnetic coupling. However, relative to α_s they have the same size. Modern MC tools hide this by just assigning them the same expansion (not numerical) value. If we are interested in the rates of new physics then other couplings in which we need to expand our computation might appear, too. In the case of standard model Higgs physics, of course, Yukawa couplings enter the game as well.

[21] Naively, the partons only have momenta in the longitudinal directions. As the proton has a size of 1 GeV we don't expect transverse momenta to be large. Nevertheless, with amount of available LHC data people are interested in transverse momentum distributions, too [42].

[22] The derivation of the DGLAP equation relies on the factorization theorem. That is why its associated scale is called factorization scale μ_F.

[23] One obtains the upper line from the lower one easily by differentiation

splitting[24] probability between the scales μ and μ_0. The second term involves exactly one splitting at the intermediate scale μ'. The primed Sudakov form factor guarantees that no splitting occurred before. Then we fold in the pdf again. We see, that if we build an iterative solution we sum an infinite tower of radiations. It is exactly the factorization theorem which allows us to do this. The integral ensures that all possible radiation histories are taken into account. Nevertheless, the pdfs are finite objects treated by the finite DGLAP equation. This is what we mean by resummation being the proper treatment of the divergent structures, which would naively produce infinitely many radiations. However, these resummed jets still could be detected. That is why we compute the cross section for $h_1 h_2 \to P + X$. The master equation describes inclusive cross sections[25]. If the final state P contains n jets (with respect to a given jet algorithm), then the cross section computation predicts the rates of events with *at least* n jets. That fact is denoted by X.

As mentioned before, scales play a fundamental role in physics. Therefor, we need a prescription to chose them for a given computation. That's where our dilemma starts. There is no such thing as a correct choice [24]. What we do instead, is to take a *reasonable* choice[26]. The scale occurring in the pdfs is called factorization scale, because it is exactly the point from which on we us the factorization theorem to resum additional radiation via the DGLAP equation. Radiation above this scale has to be encapsulated in the partonic cross section. It is clear that we can always shift one jet from one region to the other. As discussed together with the factorization theorem additional not resummed jet radiation causes large logarithms. These we need to avoid. Otherwise we would need to resum them, too. This happens for example in Higgs searches where jet vetoes play a role [43][27]. Avoiding logarithms through scale choices tells us on the other hand that we cannot fix the scale precisely, but only the order of magnitude. The result of only having logarithmic control however, is that we introduce an uncertainty in our cross section computation which depends on our scale choice.

This is not the only uncertainty we have to deal with. In general, we have to deal with the following sources of uncertainties when computing cross sections [22, 24, 25]:

- parametric uncertainties: The parameters defining the standard model need to be measured in experiment and come along with their own uncertainties. The largest influence usually comes from the strong coupling constant α_s. To estimate its impact we can vary α_s within its experimental uncertainties in our computation. For the LHC one usually finds effects of $\mathcal{O}(10\%)$.

- perturbative uncertainties: The truncation of the perturbative series automatically introduces uncertainties of $\mathcal{O}(\alpha_s)$. However, typical K-factors for QCD NLO computations range from $\mathcal{O}(1)$ to $\mathcal{O}(2)$. For large K-factors higher order computations are needed then to show that the perturbative expansion converges.

- scale uncertainties: Because we do not know what the correct scale choice is, we usually vary some central scale choice, which is given by the topology of the hard process[28], by factors of two or four. In practice, we choose the renormalization scale and the factorization scale to be the same. To test this assumption we can also vary them independently. The arbitrariness of this procedure reflects itself in the fact that the uncertainties we get, can become unreasonably huge, e.g. $\mathcal{O}(1)$ to $\mathcal{O}(100)$ for some LO computations[29]. The QCD Lagrangian does not depend on any scale and thus should neither the result of our computation[30]. The impact

[24] We call additional radiation of partons splittings, because their probability is described by splitting kernels. These kernels are computed by studying a $1 \to 2$ particle splitting configuration.

[25] Computing fixed order cross sections at a lepton collider is inclusive, too. As mentioned earlier final state partons can branch into two or more jets, too.

[26] It is a familiar fact, that the coupling runs logarithmic. The same is true for everything connected to the factorization scale. Therefor, scale choices can only be argued for in their order of magnitude.

[27] One might have the impression, that we completely switched from fixed coupling order to fixed logarithmic order. However, this is not true. We always have to combine the fixed coupling computation with the logarithmic one. Dependent on the case that might be highly non-trivial. Iterating the factorization theorem we reach σ_0 at some point, which is exactly a fixed order cross section. It is the large n limit, which we treat with resummation to compute physical probabilities.

[28] Producing a heavy particle of mass m sets the scale somewhere at $\mathcal{O}(m)$.

[29] This huge uncertainties actually appear in exclusive cross section computations, which we discuss in the next paragraph.

[30] Note: this is not a contradiction to our previous discussion about scales in physics! Physics should change depending on the scale at which we probe it. The observation of the running coupling is another good example for this fact. However, the computation of an observable quantity like a cross section should not depend on the choice

of scale variations are a remnant of our perturbative ansatz. Higher order computations can therefor strongly shrink scale uncertainties by cancellation effects. In addition resummation, which is the technique used to treat the incoming partons, also helps to shrink the scale uncertainties of observables.

- logarithmic uncertainties: Observables which suffer from large logarithms need to be resummed. As argued before this resummation is a truncated perturbative expansion with a new expansion parameter. This truncation introduces uncertainties in the same manner as the truncation of the expansion in the coupling. Treating logarithms better with higher order logarithmic computations can help to shrink scale uncertainties, too.

- numerical uncertainties: Almost always we have to do computations numerically. Computers, however, can only store finite size numbers. In addition, we need to check for the proper convergence of integrations. Usually both issues effect results with negligible impact as they can be over come by stronger computer power, i.e time.

If we are to use any theoretical computation as comparison base to data, we need to understand and control all these uncertainties. Only then we can assign statistical meaning to new physics interpretations. In addition one needs to remember that some theoretical computations are done in certain limiting cases. The factorization theorem is one example for this. Another one is the coupling of the Higgs to two gluons via a top loop, where the coupling reduces to a number in the heavy top limit [24]. Particular care has to be taken if moving away from the validity region of such assumptions. Note that independently, experiments have to deal with systematic uncertainties sourced in their understanding of the detector and, especially for rare processes, have to face limited statistics.

1.1.5 Exclusive cross sections

In contrast to the previous paragraph exclusive cross sections have a fixed number of final state objects, e.g exactly n jets. The advantage of such a prescription is, that we get statistically uncorrelated observables. In the inclusive case, for example, the three and four jet cross section are highly correlated. Therefor, exclusive cross sections open the path to safely use multi-jet observables. These are observables which are constructed from the four momenta of the observed jets. A simple example would be the transverse missing energy, where we sum all the $\vec{p}_T$ of the visible final state objects. This observable identifies not only neutrinos, but also many new physics models like dark matter candidates or invisible Higgs decays. Due to detector effects and the detectors fiducial volume jets will contribute to this observable even if there was no invisible particle involved. From an inclusive computation we could not construct this observable. A further big difference is that inclusive computations open the road to rate computations and differential distributions only. However, they do not tell us anything about the kinematic structure of a single event. The power of this knowledge is, for example, demonstrated in top tagging algorithms [44], where the substructure of a single fat jet is studied. It is therefor desirable to compute exclusive cross sections and construct the kinematic structure of single events.

The basic kinematic information is of course encoded in the matrix element. However, we already saw, that a parton coming from the hard interaction undergoes an evolution in which it will radiate additional partons. This radiation pattern is closely connected to the total number of jets observed in the event, but also to the kinematic distribution within a given jet. Given a matrix element configuration with some external partons, how can we evolve them to a full event with proper jets? For simplicity let us start with a single parton. The answer to our question lies within Eq. (1.6) and (1.7) and is known as parton shower [24, 25]. Taking the interpretation of the Sudakov form factor literally, that it is the non splitting probability between two scales, we can invert that relation and compute the scale where a splitting occurred. This way we can build a Markov chain to dress the original partons to get the full event structure. This is the method of the parton shower. Take the scale μ_F of the hard interaction and some lower cut off scale μ_0, where hadronization effects become important. This is exactly the intermediate area of the picture

of any *intermediate* parameter. What μ_F technically does, is that it separates the additional radiation computed by the matrix element from that which is resumed with the DGLAP equation. The final result should not depend on this choice.

developed in the beginning. We want to compute the intermediate scale μ' of the first branching. Once this is done we only need to iterate until we reach μ_0. We solve

$$\Delta_i\left(\mu_F, \mu'\right) = \mathcal{R}, \tag{1.8}$$

for μ', where $\mathcal{R}$ is a random number [24,25]. If μ' is outside the range of μ_0 we stop the procedure. The energy and angular distribution we then get from the corresponding splitting kernel[31]. The μ' gained that way then serves as the new μ_F for the next splitting. Using this procedure we can fill the full time evolution of the event. We get all the kinematic distributions of the partons. Once the event is *"full"* we cluster it into jets. We now have constructed an exclusive event. It is important to note that the proper procedure is much more involved then the sketch we give here. In particular we have to take care of the incoming parton line, too. There we have to fold in the pdf's to obtain the correct result [24, 25]. In addition we have to define how many partons should come from the matrix element. If we compute up to five partons in the matrix element a three parton matrix element with two branchings could describe the same state. This obviously results in a double counting problem. Matching algorithms like CKKW [45] and MLM [46] have been developed to take care of this issue. Algorithms like these are implemented in state of the art Monte Carlo methods. Sherpa [47], MadGraph [48], and Alpgen [49] together with Pythia [50] and Herwig [51] use the concepts described here, but also a lot of additional physics knowledge, to simulate full events. These we can compare to the data at the LHC, but also to think about new analysis strategies. These Monte Carlo codes are the work horses of theoretical and experimental particle physics and also represent the link between the two.

In the previous paragraph we use the Sudakov form factor to develop the method of the parton shower, which is implemented in modern simulation tools. There is, however, also another path starting from Eq. (1.6). The DGLAP equation does not tell us what the pdf actually is. We put that information in by hand and interpret it as the momentum fraction distribution within the proton. From the structure of the equation we could put in any parton distribution we want. We use this insight to compute the number distribution of the partons. The method is known as generating functional formalism [25]. In contrast to the previous described parton shower we do not get a full kinematic event distribution, but on the other hand we get an analytic prescription of exclusive event rates. This can be of great advantage if we want to study jet radiation at the LHC. Analytic equations help us to understand better the sources of some of the effects we observe. Given the fraction P_n of events with n jets for a given hard matrix element σ_0 the generating function is then defined as

$$\Phi(u) = \sum_n u^n P_n. \tag{1.9}$$

We get the rates back by differentiation[32] with respect to u at $u = 0$.

$$P_n = \frac{1}{n!}\frac{d^n}{du^n}\Phi(u)\bigg|_{u=0}. \tag{1.10}$$

It is clear, that the number of possible branchings depends on the distance between the scales μ_F and μ_0. If these scales coincide there is no space for branching and we have $\Phi(u, \mu_0) = u$. If we move the scales apart the generating functions have to obey the DGLAP equation. However, this time we know the initial condition.

$$\Phi_i(u, \mu_F) = \Delta_i(\mu_F, \mu_0)\Phi_i(u, \mu_0) + \sum_{i\to jk}\int_{\mu_0}^{\mu_F}\frac{d\mu'^2}{\mu'^2}\Delta_i\left(\mu_F, \mu'\right)\int\frac{dz}{z}\frac{\alpha_s}{2\pi}P_{i\to jk}\Phi_j(u, \mu')\Phi_k(u, \mu'). \tag{1.11}$$

Using this equation we can compute the rates for exclusive n jet cross sections for a given hard matrix element by multiplying a factor of $\Phi_i(u, \mu_F)$ for each parton i. In contrast to the numerical methods using matching and the parton shower, we cannot naively compute the kinematic distribution or use several matrix elements to describe more jets more precisely. Note, that this

[31] The splitting kernels are averaged over angles, thus for a correct description we include additional knowledge about polarization correlations.

[32] In general u is a function and we need a functional derivative. Therefor, the name of the method.

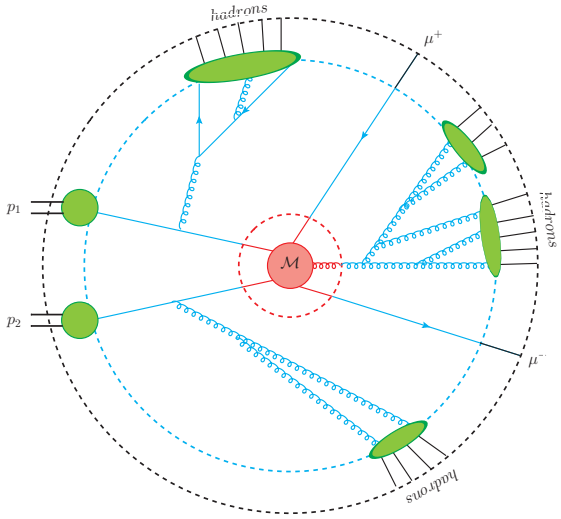

Figure 1.4: Phenomenological illustration of a Z plus jets event (compare Fig. 1.2). The colors are connected to the different concepts we use to describe the corresponding part of the evolution of the particles. Red: short range interaction, Blue: partonic evolution, Green: hadronization, Black detector interaction. The circles are interpreted as radii of equal time distance.

time we have two generating functions after the splitting kernel. This is because the parton gets destroyed by a splitting and replaced by its daughters. The advantage we obtain from knowing the initial condition is tremendous. We can write down the integral equation in a closed exponentiated from [25].

$$\Phi_i(u,\mu_F) = u \exp\left(\int_{\mu_0}^{\mu_F} d\mu'^2 \sum_{i\to jk} \Gamma_{i\to jk}(\mu_F,\mu')\left[\frac{\Phi_j(u,\mu')\Phi_k(u,\mu')}{\Phi_i(u,\mu')} - 1\right]\right), \tag{1.12}$$

where $\Gamma_{i\to jk}(\mu_F,\mu')$ is the integrated splitting kernel $P_{i\to jk}(z)$. The integration boundary depends on the actual scale μ_F. The integration is performed exactly in the logarithmic expansion mentioned earlier. At NLL we have [25][33]

$$\begin{aligned}
\Gamma_{g\to gg}(\mu_F,\mu') &= C_A \frac{\alpha_s}{2\pi\mu'^2}\left(\log\frac{\mu_F^2}{\mu'^2} - \frac{11}{6}\right),\\
\Gamma_{g\to q\bar{q}}(\mu_F,\mu') &= T_R \frac{\alpha_s}{2\pi\mu'^2}\left(\frac{n_f}{2}\right),\\
\Gamma_{q\to qg}(\mu_F,\mu') &= C_F \frac{\alpha_s}{2\pi\mu'^2}\left(\log\frac{\mu_F^2}{\mu'^2} - \frac{3}{2}\right).
\end{aligned} \tag{1.13}$$

1.1.6 The full picture

We have collected all the information to understand the basics of collisions at the LHC. In Fig. 1.4 we show the concepts introduced before and at which time scale they live to put everything in a closed illustration. At the heart of each interaction we have the short range physics, which are described with matrix elements computed in perturbative QFT (red region). At the LHC each matrix element has at least two external partons. Moving to larger time scales each parton evolves and generates a complicated radiation pattern, which we describe with the parton shower or the generating functional formalism (blue region). Two of the resulting parton lines need to be connected to the proton and are thus weighted with pdf factors (green disks at p_1 and p_2). All other partons are seeds for hadronization (green blobs). The hadrons then decay. Together with

[33] Remember that C_A, C_F and T_R are color factors. The number of flavors n_f in the standard model is six. However, as the top and b quarks are heavy we usually deal with four or five flavors in applications.

the other stable elementary particles[34] they leave a distinct pattern in the detector enabling us to identify them (black circle).

There are more effects then these described here, which effect the signatures we detect in the experiments. There is the possibility that more than one parton per proton interacts, known as multiple parton interaction (MPI). In addition we have to face the fact that not only one proton per bunch crossing produces scattered particles. This effect is known as pile up (PU). For certain applications jets do offer the possibility to deal and correct for these so called soft effects via trimming methods [52]. Hadronization needs to be modeled and cannot be derived from first principles [25]. We get model dependent effects therefor. This is also true for the rest of the proton, the so called beam remnants. Modern MC simulations include all these effects. However, they all will be of no concern to us regarding our analytic ccomputations.

1.2 The power and trouble of multi jet observables

Let us now put this knowledge to use. We study the jets plus missing energy channel. Missing energy is a generic dark matter candidate signature. In addition, we assume these dark matter particles to be produced from the decay of some colored new physics. This will produce additional jets, whose structure differs from the simple evolution of QCD partons described before. We show how understanding jet radiation enables us to construct fairly model independent[35] search strategies. The focus in our study is on the standard model background as a tool to identify new physics. Therefor, we discuss the theoretical uncertainties, which confront our method, in detail. This will be exactly the scale and parametric uncertainties mentioned in the context of cross section computations. Irrespective of the size of the uncertainties we find that standard model backgrounds obey so called staircase scaling patterns. These will be the seed of our further theoretical considerations.

The rest of this section has already been published in [27] together with Christoph Englert, Steffen Schumann and Tilman Plehn.

1.2.1 Jets with missing energy

Missing transverse energy is a general signature for dark matter related new physics at hadron colliders [53]. It has a long history at the Tevatron and to date gives the strongest bounds on squark and gluino masses in supersymmetric extensions of the Standard Model. At the LHC the first new exclusion limits for squarks and gluinos have recently appeared, in the CMSSM toy model as well as in a more general setup [54–56]. All of these analyses are based on jets plus missing energy including a lepton veto which constitutes the most generic search strategy for strongly interacting new particles decaying into a weakly or super-weakly interacting dark matter candidate [53, 57].

While the first results are based on very inclusive cuts, following the ATLAS [58] and CMS [59] documentations we expect more specific analyses to appear soon. The reason is that in their current form the analyses can and should be optimized for specific new physics mass spectra. More specialized analyses for jets plus missing energy rely on a missing transverse momentum cut and on a certain number of staggered jet transverse momentum cuts [58, 59]. Unfortunately, they are therefore hard to adapt to modified mass spectra and by definition show a poor performance for not optimized model parameters. In addition, they are counting experiments in certain phase-space regions, which means that for any additional information on the physics behind an anomaly we have to wait for a dedicated analysis.

A major problem of searches for new physics in pure QCD plus missing energy final states is the prediction of background distributions. Aside from the improved signal-to-background ratio this is one of the reasons why applying fairly restrictive cuts on the number of jets and on their transverse momentum is a promising strategy. Such cuts relieve us from having to understand the complete p_T spectra [60] of general exclusive or inclusive n_{jets}-jet events at the LHC. Experimentally, however, we should by now be in a position to simulate these distributions using the CKKW [45, 61] or

[34] In principle leptons would undergo an evolution, too. However due to the smallness of the electromagnetic coupling α we ignore this effect.

[35] The assumption of a strong production and decay jets is already a constraint. However, what we mean is, that we do not need to test for particular model parameters.

MLM [46] matching methods implemented in SHERPA [47], ALPGEN [49], or MADEVENT [48]. The different approaches have been compared in some detail, for example for W+jets production [62, 63]. What is still missing is a systematic study of theory uncertainties in multi-jet background simulations for top quark analyses and new physics searches, i.e. including large jet multiplicities down to intermediate jet transverse momenta, but reflecting a well defined hard scale given by the signal process. Motivated by theoretical and statistical considerations we define all observables as exclusive, specifically in the number of jets.

In this paper we establish a proper simulation of multi-jet processes and estimate these theory uncertainties, with a focus on the question what actually constitutes the theory error. This way LHC data in control regions can be used to understand very generic scaling features (*staircase scaling*[36]) which have already been observed in data [65, 66] and which we can extend based on appropriate Monte Carlo studies. This staircase scaling we can reproduce and study using QCD Monte Carlo simulations, including different hard processes and the effects of cuts. Combining these simulations with LHC data should give us a quantitative handle on multi-jet rates in many applications.

Moreover, we can use our knowledge about the exclusive n_{jets} distributions to predict other notoriously difficult multi-jet observables. So once we understand the uncertainties on the multi-jet spectra we turn to the effective mass. In its many incarnations it either includes the leading jet or it does not and is either limited to four jets or any other number of jets [67]. Obviously, any specific definition of this mass variable increases its sensitivity to theory uncertainties. We study the most generic definition of the effective mass including *all jets* visible above a transverse momentum threshold. The theory uncertainties of this observable can be closely linked to the jet-multiplicity distribution. Using the scaling properties of the exclusive jet multiplicities we can strongly reduce the theory uncertainty in the effective-mass spectrum in a consistent manner. The same should be true for other variables which we can use to extract new physics from jet dominated backgrounds.

Similar questions are currently being asked to control regions in a purely data-driven approach. However, the conversion from background regions into the signal region either by shifting the kinematic regime or by changing the hard core processes off which we radiate jets requires a good understanding for example of the effects of background rejection cuts and of background sculpting features in the definition of these observables. These effects we can reliably estimate in an appropriate Monte Carlo study and then combine for example with an over-all normalization from data.

Finally, we suggest an analysis strategy which on the one hand makes maximum use of the jet patterns and on the other hand does not require any tuning of cuts. The only ingredient of our analysis which does not involve jets is a missing transverse energy cut to reduce pure QCD backgrounds and an isolated lepton veto against W+jets backgrounds. To reduce both of them to a manageable level we require

$$\not{E}_T > 100 \text{ GeV} \qquad \text{and a lepton veto if} \quad p_{T,\ell} > 20 \text{ GeV}, \; |y_\ell| < 2.5 \tag{1.14}$$

as the basic and only electroweak cuts to reduce the QCD background. The exact numbers are not very dependent on the details of the model as long as the new physics sector provides a WIMP dark matter candidate. To account for fake missing energy from QCD jets we apply an additional factor of 1/500 for pure QCD and hadronic top-quark final states. This rough fake rate we estimate from Ref. [54]. It provides us with a rather conservative estimate compared e.g. to Ref. [59].

After these very generic acceptance cuts a two-dimensional correlation of the effective mass vs the exclusive jet multiplicity is the appropriate distribution to extract limits on strongly interacting new physics or in the case of an excess study the mass scale as well as the color charge of the new states. Because all our observables are defined jet-exclusively we can to a good approximation study this two-dimensional distribution using a log-likelihood shape analysis. The contributions of different regions in the n_{jets}- m_{eff} space to the binned log-likelihood automatically focus on the correct phase-space region and are readily available for improved analyses as well as theoretical interpretation.

1.2.2 Jet number scaling

To separate new physics events from a QCD sample after some very basic cuts we have to understand the number of jets and their energy or p_T spectra. This will allow us to exploit two

[36] Staircase scaling for jet rates is often referred to as Berends scaling. However, to our best knowledge it was first introduced and discussed by the authors of Ref. [64].

distributions: the number of jets observed (n_{jets}) and their effective mass (m_{eff}), where the definition of the latter usually requires us to define the number and hardness of the jets included in its construction. Our maximally inclusive approach means that aside from the fiducial volume of the detectors all we fix is the algorithmic jet definition to count a jet towards each of the two measurements. Throughout this paper we define jets using the anti-k_T algorithm [68] in FastJet [38] with a resolution $R_{\text{anti}-k_T} = 0.4$ and then require

$$p_{T,j} > \; p_{T,\text{min}} = 50 \;\; \text{GeV} \qquad\qquad \text{and} \qquad\qquad |y_j| < 4.5 \; . \tag{1.15}$$

This defines which jets are counted towards n_{jets} as well as the m_{eff} distribution. Given $p_{T,\text{min}}$ we can then evaluate the 2-dimensional n_{jets} vs m_{eff} plane in Section 1.2.5 using a binned log-likelihood approach.

Before we can use the n_{jets} distribution to extract new physics in the jets plus missing energy sample at the LHC we need to show that we understand this distribution in detail. Obviously, the overall normalization of this distribution is not critical. For any kind of new physics not completely ruled out by the Tevatron experiments the two jet and three jet bins are practically signal free. So the question becomes: what can we say about the shape of $d\sigma/d\, n_{\text{jets}}$.

For W+jets events this kind of distribution has been studied, even at the LHC [66]. We observe the *staircase scaling* [64, 69], an exponential drop in the inclusive n_{jets} rates with constant ratios $\hat{\sigma}_{n+1}/\hat{\sigma}_n$. The numerical value of this ratio is obviously strongly dependent on $p_{T,\text{min}}$. The original staircase scaling describes inclusive jet rates, i.e. it uses $\hat{\sigma}_n$ including all events with at least n jets fulfilling Eq.(1.15). In the light of recent advances in QCD and because our likelihood analysis should be based on independent bins we define the scaling in terms of exclusive jet rates, i.e. counting only events with exactly n jets fulfilling Eq.(1.15) towards σ_n. This preserves the normalization of the n_{jets} histogram as $\sigma_{\text{tot}} = \sum_n \sigma_n$ and makes it possible to add the bins in the computation of the log-likelihood. It is interesting to note that staircase scaling defined either way implies staircase scaling using the other definition, and that the jet-production ratios of the two approaches are identical. If we define the universal exclusive staircase-scaling factor as

$$R \equiv R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} \; , \tag{1.16}$$

we find for the usual inclusive scaling denoted by a hat over all parameters

$$\hat{R} \equiv \frac{\hat{\sigma}_{n+1}}{\hat{\sigma}_n} = \frac{\sigma_{n+1} \sum_{j=0}^{\infty} R^j}{\sigma_n + \sigma_{n+1} \sum_{j=0}^{\infty} R^j} = \frac{R\sigma_n}{(1-R)\sigma_n + R\sigma_n} = R \; . \tag{1.17}$$

The same relation we find when we include a finite upper limit to the number of jets in the sum over j. Note, however, that this argument only holds for a strict staircase scaling where the ratio $R_{(n+1)/n}$ does not depend on the number of jets n. For our analysis this means that we can use the staircase scaling for a statistical analysis of the $d\sigma/d\, n_{\text{jets}}$ distribution either in its inclusive or in its exclusive version, the latter based on independent n_{jets} bins.

In this section we will show that (1) such a scaling exists not only for W/Z+jets but also for pure QCD events and (2) we can reliably estimate the scaling factor and possible deviations from it from theory. A purely data-driven background analysis of this distribution might be possible and should be combined with our results. For example, we can one by one remove the missing energy cut and the lepton veto in Eq.(1.14) which gives us background dominated event samples to a reasonably large number of hard jets. Adding the background rejection cuts will then have an impact on the scaling, which we can estimate reliably. For the signal hypothesis we have to entirely rely on QCD predictions.

As a starting point we discuss the established staircase scaling in W+jets production. The behavior of Z+jets is exactly the same. For our analysis we produce Ckkw-matched [61] background samples for W+jets (to 5 ME jets), Z+jets (to 5 ME jets), $t\bar{t}$+jets (to 2 ME jets), and QCD jets (to 6 ME jets) with Sherpa-1.2.3 [47]. Higher order corrections to the inclusive scaling we expect to, if anything, improve the assumption of a constant jet ratio $\hat{R}$ for example in W+jets production [70].

In the left panel of Figure 1.5 we show the exclusive n_{jets} distribution for the LHC running at 7 TeV center of mass energy. To increase our statistics to large enough values of n_{jets} we do not

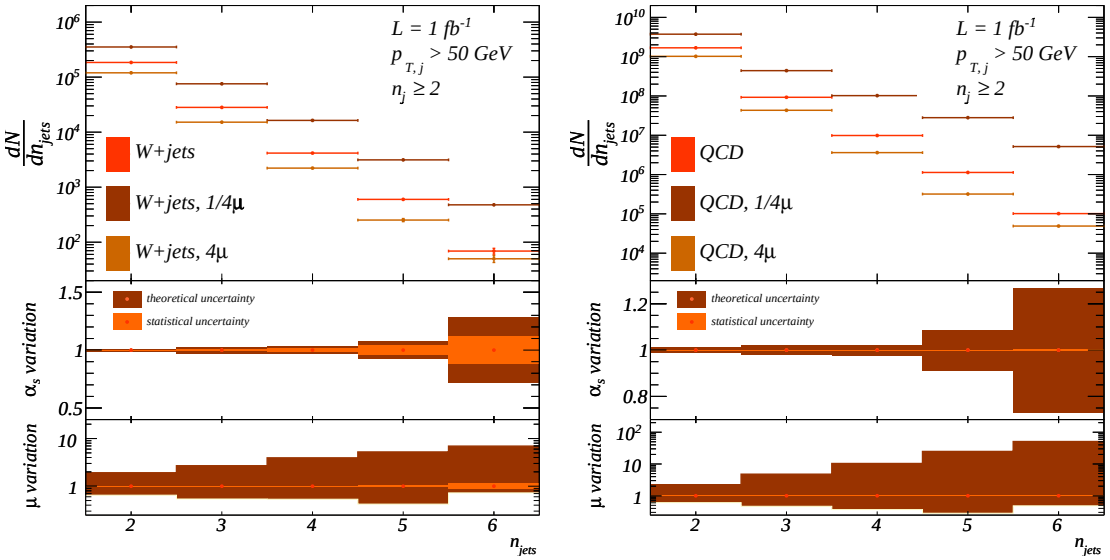

Figure 1.5: Exclusive $d\sigma/d\ n_{\text{jets}}$ distribution for W+jets (left) and QCD jets production at the LHC. Only the jet cuts given in Eq.(1.15) are applied, neither a $\not{E}_T$ cut nor a lepton veto is imposed. The second panel shows the parametric uncertainty due to a consistent change of $\alpha_s(m_Z)$ between 0.114 and 0.122. The third panel shows the reach of a consistent scale factor treatment which can be experimentally determined and should not be considered a theory uncertainty.

apply the selection cuts Eq.(1.14) in this first step. We already see that we can qualitatively fit a line through the central points on a logarithmic axis for each set of input parameters.

Before we quantitatively evaluate this scaling we need to consider the uncertainties associated with our simulation. This is crucial if we want to use the n_{jets} scaling as a background estimate for new physics searches in QCD final states. There are two distinct sources of uncertainty in our simulation. First, there exists a parametric uncertainty, namely the input value of $\alpha_s(m_Z)$ or some other reference scale. To address this, we consistently evaluate the parton densities around the central NLO value $\alpha_s(m_Z) = 0.118$ inside a window $0.114 - 0.122$ [71] and keep this value for all other appearances of the strong coupling in our matrix-element plus parton-shower Monte-Carlo simulations. In Figure 1.5 we see that the resulting error bar on the $d\sigma/d\ n_{\text{jets}}$ increases with the number of jets, but stays below 30% even for the radiation of six jets. For luminosities around 1 fb^{-1} the error on α_s is roughly of the same order as the experimental statistical error. Systematic errors we do not consider, even though they will at some point dominate over the statistical errors. After any kind of realistic background rejection the combined experimental error will exceed the parametric α_s uncertainty.

The reason why we cannot use staircase scaling in W+jets to measure α_s is a second source of QCD uncertainty: aside from the parametric α_s error band, an actually free parameter in our QCD simulation is a common scaling factor μ/μ_0 in all appearances of the factorization and renormalization scales, including the starting scale of the parton shower. Identifying all scales follows the experimental extraction of the parton densities and α_s in a simultaneous fit. The interpretation of DGLAP splitting in terms of large logarithms tells us that the factorization and renormalization scales have to be identified with the transverse momentum of the radiated jets. By definition, such leading-logarithm considerations leave open the proportionality factor in the relation $\mu \propto |p_{T,j}|$. Any constant factor can be separated from the dangerous logarithm as a non-leading constant value.

Because this constant cannot be derived from first principles we vary it in the range $\mu/\mu_0 = 1/4 - 4$ and show the numerical result in Figure 1.5. As expected, the variation of the jet rates with this scaling parameter is huge — much larger than the experimental uncertainties we expect from the LHC and which we know from the Tevatron. In Figure 1.5 we can first of all check that introducing such a scaling factor does not seriously impact the observed staircase scaling. Counting such a constant towards the theory uncertainty is questionable if we can determine it experimentally. For example for SHERPA we know from Tevatron that the scaling factor should essentially be unity [72], which in the spirit of Monte-Carlo tuning means that for example in

channel (cuts)	$R_{2/1}$	$R_{3/2}$	$R_{4/3}$	$R_{5/4}$	$R_{6/5}$	$R_{7/6}$	R_0	$\frac{dR}{d\ n_{\text{jets}}}$
	SHERPA simulation						linear fit	
W+jets ($p_{T,j} > 50$ GeV)	0.1931(3)	0.1494(5)	0.157(1)	0.138(3)	0.115(8)	0.09(2)	0.150(1)	−0.001(1)
W+jets (+ lepton veto)	0.2290(4)	0.1494(7)	0.164(2)	0.139(4)	0.12(1)	0.09(2)	0.149(1)	−0.002(1)
W+jets (+ $\not{E}_T > 100$ GeV)	0.252(1)	0.224(2)	0.190(5)	0.16(1)	0.15(2)	0.09(4)	0.239(3)	−0.032(3)
Z+jets ($p_{T,j} > 50$ GeV)	0.1463(2)	0.1504(6)	0.147(1)	0.138(4)	0.123(9)	0.07(2)	0.154(1)	−0.006(1)
Z+jets (+ $\not{E}_T > 100$ GeV)	0.2251(6)	0.185(1)	0.166(3)	0.154(6)	0.14(1)	0.08(3)	0.193(2)	−0.018(2)
QCD jets ($p_{T,j} > 50$ GeV)	—	0.0552(1)	0.1074(5)	0.106(1)	0.125(5)	0.12(1)	0.105(2)	0.001(1)
$(t\bar{t})_{hh}$+jets ($p_{T,j} > 50$ GeV)	3.69(9)	1.26(2)	0.67(1)	0.366(9)	0.24(1)	0.15(5)		
$(t\bar{t})_{\ell\ell/h}$+jets ($p_{T,j} > 50$ GeV)	1.96(2)	0.851(7)	0.465(5)	0.260(5)	0.168(8)	0.12(2)		
$(t\bar{t})_{\ell\ell/h}$+jets (+ lepton veto)	1.75(2)	0.765(10)	0.391(7)	0.228(8)	0.14(1)	0.12(3)		
$(t\bar{t})_{\ell\ell/h}$+jets (+ $\not{E}_T > 100$ GeV)	1.60(5)	0.83(2)	0.49(2)	0.25(2)	0.15(2)	0.19(7)		

Table 1.2: Jet ratios for all Standard Model channels, including (semi-)leptonic and hadronic top pairs for the central scale choice $\mu = \mu_0$. The quoted errors are statistical errors from the Monte Carlo simulation. The numbers correspond to the curves shown in Figures 1.5 and 1.6.

SHERPA the naive default parameter choice comes out as correctly describing the data. Of course, this does not have to be true for other simulation tools. An interesting question to ask once we have access to it at the LHC would be if this *per se* free parameter really is the same for different channels, like W/Z+jets and QCD jets.

In the right panel of Figure 1.5 we show the same distributions for pure QCD jet production. Again, not applying the cuts in Eq.(1.14) we observe staircase scaling, however, with some caveats for the two and three jet bins. This is related to the definition of the hard process. As expected, the scale factor μ/μ_0 has very large impact not on the existence of a staircase scaling but on the jet ratio R. The parametric uncertainty due to the error bar on $\alpha_s(m_Z)$ is again small once we vary the strong coupling consistently everywhere, staying below 30% for up to six jets. The parametric uncertainty for the pure QCD case and the W+jets case is clearly very similar. The scale factor variation $\mu/\mu_0 = 1/4 - 4$ gives an even larger band of possible ratios of cross sections, to be contrasted with a reduced statistical uncertainty compared to the W+jets case. Our argument that this over-all scale factor should be determined experimentally is therefore even more applicable for the QCD case. To date such an analysis does not exist, so while in the following we will use unity as the appropriate scale factor for SHERPA this needs to be verified experimentally.

Once we understand the size of theory uncertainties for the exclusive $d\sigma/d\ n_{\text{jets}}$ distribution we need to quantify the quality of the observed staircase scaling. Since the quantitative outcome will depend on the background rejection cuts we apply, we study the scaling without the cuts shown in Eq.(1.14), after the lepton veto only, and including the lepton veto as well as the missing energy cut. Starting from the individual $R_{(n+1)/n}$ values we fit a line through all relevant data points, as

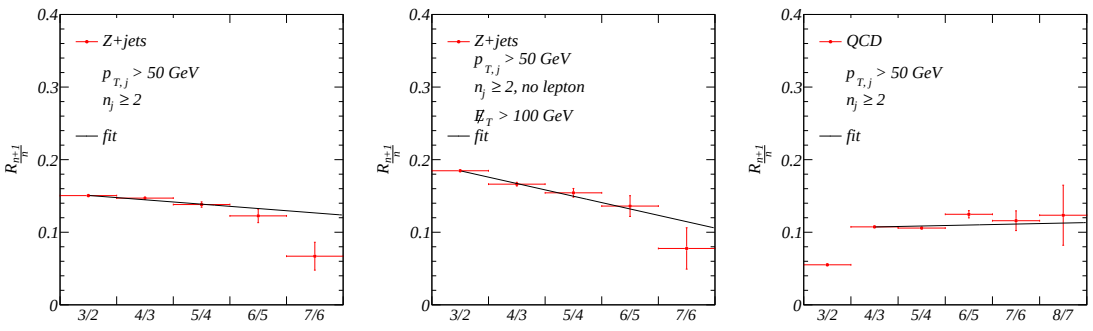

Figure 1.6: Jet ratios for Z+jets production (without and with the $\not{E}_T > 100$ GeV cut) and QCD jets production, corresponding to the numbers listed in Table 1.2. The error bars indicate the remaining Monte Carlo uncertainty in our simulation.

a function of n_{jets}

$$R(\ n_{\text{jets}}) = R_0 + \frac{dR}{d\ n_{\text{jets}}}\ \ n_{\text{jets}}\ , \tag{1.18}$$

and determine the slope to compare it to our prediction $dR/d\ n_{\text{jets}} = 0$.

In Table 1.2 we list the exclusive jet ratios as shown in Figure 1.5. For the W/Z+jets case we see that the radiation of one compared to a second jet off the Drell-Yan process $R_{2/1}$ does not show this scaling. The reason for this specific feature is the definition of the hard core process alluded to before. To generate the relatively hard jets and the large missing energy mimicking the signal events we need to at least consider W/Z+1 jet as the core process. In addition, we do not take into account any separation criterion between the first jet and the gauge boson, which means we treat σ_1 different from all other σ_n. In Table 1.2 we see that we are lucky for the Z+jets case, but we are not for the W+jets case. The tricky definition of the hard process σ_1 as the base of additional jet radiation suggests that we start our staircase scaling analysis with $R_{3/2}$.

The statistical uncertainties which we show in Table 1.2 and which enter the fit of the slope as defined in Eq.(1.18) always increase towards larger jet multiplicity. This is an effect of the way we simulate these events which completely corresponds to an experimental analysis. If we generate (or measure) all n_{jets} bins in parallel the first bin will always have by far the smallest error. Therefore, it determines the constant scaling factor R_0 in our fit as well as in a measurement. For larger values of n_{jets} we become statistics dominated, which means that Monte Carlo simulations can extend the reach of actual measurements at any given point in time towards larger jet multiplicities. This is the phase space region in which we need to provide new physics searches at the LHC with accurate background estimates.

Some of the rows listed in Table 1.2 we also depict in Figure 1.6. For electroweak gauge boson production we see that without any cuts W and Z production show the same scaling parameter R_0 as well as a small negative slope. Within errors the staircase scaling holds to six and possibly seven jets, even though we see a slight slope developing towards larger numbers of jets. This is a phase space effect which is expected once we start probing gluon parton densities and their sharp drop towards larger parton momentum fractions and which is well modeled by our computation.

Adding the lepton veto does not change the staircase scaling at all. This means that forcing the W boson to decay into one fairly soft lepton and a harder neutrino does not affect the behavior of the recoiling jets. Adding a significant $\not{E}_T$ cut, on the other hand, has a measurable effect on the jet ratios as well as on the slope. For experimental applications of this scaling, however, it is important to note that the phase space effects for large n_{jets} as well as the effect of kinematic cuts are completely described by our simulation.

For pure QCD events we find a remarkable agreement with the staircase scaling hypothesis, which seems to be supported by recent LHC analyses [73]. The definition of the hard core process is somewhat problematic since there exists no inherent hard scale in the $2 \to 2$ process and the infrared behavior of s-channel and t-channel diagrams is very different. Therefore, we define σ_3 as the starting point of our analysis and $R_{4/3}$ as the first relevant cross section ratio. Table 1.2 and Figure 1.6 show that the ratios $R_{(n+1)/n}$ are essentially constant to eight jets. The slope within statistical uncertainties is, in contrast to W/Z production, fully compatible with zero. The central R_0 values for W/Z+jets and QCD jets production are slightly different, which is expected by the different core processes and by the different background rejection cuts.

1.2.3 Decay jets vs jet radiation

In contrast to the QCD and gauge-boson background n_{jets} distributions from heavy particles decaying to jets include two sources of jets: first, there are decay jets, which dependent on the spectrum might or might not be hard enough to stick out. Second, there is QCD jet radiation, which for heavy states will generically be relatively hard and dominated by collinear splitting in the initial state [74, 75], leading to a non-zero maximum value of the number of expected initial-state radiation jets [53, 76]. Due to the hard scale of new-physics processes on the one hand and because we need to simulate supersymmetric decays inclusively we best generate the new-physics events with Herwig++-v2.4.2 [77] and normalize the cross sections with Prospino2.1 [78]. All supersymmetric mass spectra we generate with SoftSusy [79] using the SLHA output format [80] and use Sdecay [81] to calculate the leading-order branching ratios. We check the jet-radiation results from the Herwig++ shower with Mlm merging implemented in MadEvent [75], using

PYTHIA [50] for parton showering and hadronization. As expected, the two simulations agree well within their uncertainties.

The question for heavy-particle production is how universal its n_{jets} distributions are when we consider Standard Model as well as new-physics particles with different masses and color charges, like top quarks, squarks and gluinos. In Figure 1.7 we first show the n_{jets} distributions for (semi-)leptonic and hadronic top-pair production. We see how all unsubtracted distributions show maxima away from $n_{\text{jets}} = 0$, driven by the presence of decay jets plus relatively hard jet radiation. In addition, they do not show a staircase scaling at large jet multiplicities. Because the particles produced in the hard process have non-negligible masses even compared to the hadronic center-of-mass energy the phase-space suppression for example due to rapidly dropping gluon densities kicks in immediately and bends the otherwise exponential fall-off.

In the Standard Model we can fit the (semi-)leptonic and purely hadronic top-pair distributions for all jets fulfilling Eq.(1.15) to the function

$$\frac{d \log \sigma(\ n_{\text{jets}})}{d\ n_{\text{jets}}} = -b\ \frac{n_{\text{jets}}^2 - a_1\ n_{\text{jets}} + a_0^2}{n_{\text{jets}}}\ . \tag{1.19}$$

The two relevant fit parameters for the normalized distributions shown in Figure 1.7 correspond to the maximum at $n_{\text{jets}} = a_0$, and the (staircase) scaling parameter for QCD jet radiation at large n_{jets} given by $R = \exp(-b)$. Because we do not include higher suppression terms towards large n_{jets} we stop the fit at the endpoints of the curves shown in Figure 1.7.

In Table 1.3 we list the best fit values for these parameters for both top decays. We immediately see more quantitatively than in Figure 1.7 that for example hadronically decaying top pairs on average include not even one more jet than the (semi-)leptonic sample. Typically only one of two jets from the W decay is accounted for because of the cutoff at $p_{T,\text{min}} = 50$ GeV. Comparing this value to the W mass it is likely that one of the two W decay jets gets boosted above $p_{T,\text{min}}$, but the other one stays below. In contrast, the Jacobian peak of the b-quark energy from the top decay lies above $p_{T,\text{min}}$. Going back to Table 1.3 this means that for top pairs the most likely number of radiated jets is zero, closely followed by one jet emission [76].

For squarks and gluinos the features we see in top-pair production become more pronounced and the jet multiplicity reflects the color charge of the produced particles. As a reference point in supersymmetric parameter space we consider reasonably low mass gluinos and squarks in the SPS1a benchmark scenario [82], with $m_{\tilde{g}} = 608$ GeV and typical light-flavor squarks around $m_{\tilde{q}} \sim 558$ GeV. The new LHC exclusion limits are right at the edge of excluding this standard parameter choice[37]. Because the gluino cannot decay to a gluon it requires two quarks to get rid of its color charge. Squark pairs, including squark-antisquark production, predict two hard decay jets

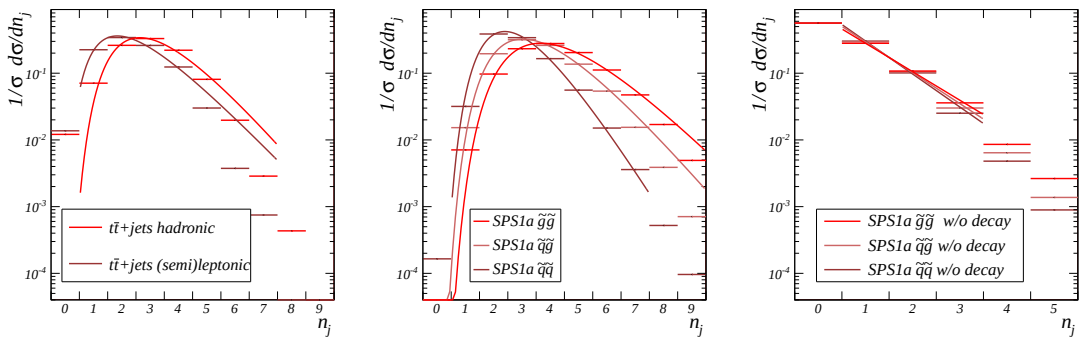

Figure 1.7: Normalized exclusive $d\sigma/d\ n_{\text{jets}}$ distributions for top pairs (left) and supersymmetric particle production. For the latter we show all decay jets plus QCD jet radiation (center) as well as QCD jet radiation only (right). Jets are counted once they fulfill Eq.(1.15).

[37] Due to the presentation of the LHC results in the m_0 vs $m_{1/2}$ plane it is also not possible to precisely read off the actual limits in terms of physics mass parameters. Moreover, since squark and gluino masses are both mostly driven by $m_{1/2}$, there does not exist a mapping of the m_0-$m_{1/2}$ plane into the squark-gluino mass plane. Models with significantly heavier gluinos than quarks are excluded in CMSSM searches.

	$(t\bar{t})_{hh}$	$(t\bar{t})_{\ell\ell/h}$	$\tilde{q}\tilde{q}$	$\tilde{q}\tilde{g}$	$\tilde{g}\tilde{g}$	SUSY	$\tilde{q}\tilde{q}$	$\tilde{q}\tilde{g}$	$\tilde{g}\tilde{g}$
	full		full				jet radiation		
a_0	3.13	2.34	2.89	3.53	4.16	3.15	n.a.	n.a.	n.a
a_1	5.41	3.73	5.28	6.16	7.15	5.48	0.45	0.36	0.21
b	1.25	1.07	1.71	1.25	1.09	1.27	1.14	1.07	0.98

Table 1.3: Parameters defined in Eq.(1.19) and extracted from the unsubtracted distributions shown in Figure 1.7. The parameter a_0 corresponds to the position of the maximum while b captures the approximate scaling at larger n_{jets}. The combined supersymmetric result is based on the appropriately weighted event samples for squarks and gluinos.

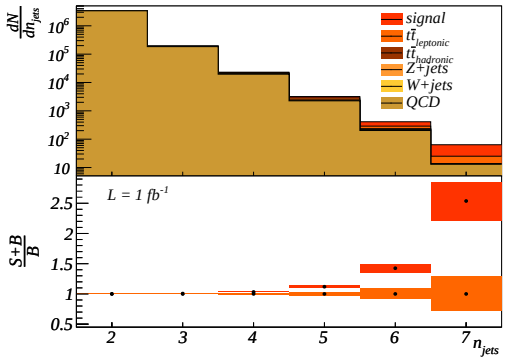

Figure 1.8: Exclusive n_{jets} distribution for all considered Standard-Model backgrounds and the SPS1a signal for supersymmetry. We present the results for an LHC center-of-mass energy of 7 TeV with an integrated luminosity of 1 fb^{-1} and after the cuts specified in Eqs. (1.14) and (1.15).

plus some QCD radiation and sub-leading decay jets. In Table 1.3 we see that for this production channel the maximum of a continuous n_{jets} distribution indeed resides almost at $n_{\text{jets}} = 3$. For associated squark-gluino and gluino-pair production the number of jets increases by almost one, corresponding to the second gluino-decay jet which not in all cases is hard enough to appear after requiring $p_{T,\text{min}} = 50$ GeV. The jet multiplicity of the entire supersymmetric sample is close to the average for squark pair production and squark-gluino production which reflects the hierarchy in cross sections of the three processes [78].

Breaking down the supersymmetric signal into individual production processes we can examine the distinct radiation patterns. Gluino pairs radiate significantly more than associated production or squark pairs, which is reflected in the right columns of Table 1.3: $b(\tilde{g}\tilde{g}) < b(\tilde{q}\tilde{g}) < b(\tilde{q}\tilde{q})$. The scaling parameter $R = \exp(-b)$ is consistently larger than for the background samples in Table 1.2. For example for the jet radiation off squark pair production we find $R \approx 0.32$. Moreover, in Figure 1.7 we see that the jet rates for QCD radiation drop off even faster for large multiplicities. This means that there definitely does not exist any staircase scaling behavior for heavy particle pair production above a threshold of 1 TeV at the LHC with a hadronic center-of-mass energy of 7 TeV. This phase space argument should not be mixed with the fact that the hard scale of such processes and with it the logarithmic enhancement for collinear radiation is large, i.e. the validity of the collinear approximation extends to larger values of $p_{T,j}$.

Finally, in Figure 1.8 we show the n_{jets} distribution for the supersymmetric signal assuming the SPS1a parameter point and the various Standard Model backgrounds. We apply the background rejection cuts specified in Eqs.(1.14) and (1.15). The variation in shape when including the signal events is statistically significant and appears as an excess of high jet-multiplicity events for $n_{\text{jets}} > 5$. The associated statistical significance we compute in Section 1.2.5.

1.2.4 Effective mass

Before we turn to exploit the number of jets to extract a new physics signal at the LHC an obvious question is if we can make use of our understanding of the n_{jets} distribution looking at other observables in multi-jet final states. More specifically, we will use the measured scale parameter μ/μ_0 shown in Figure 1.5 to reliably predict observables, which, based on traditional QCD simulations, show an overwhelming theory uncertainty. A classic observable in this respect

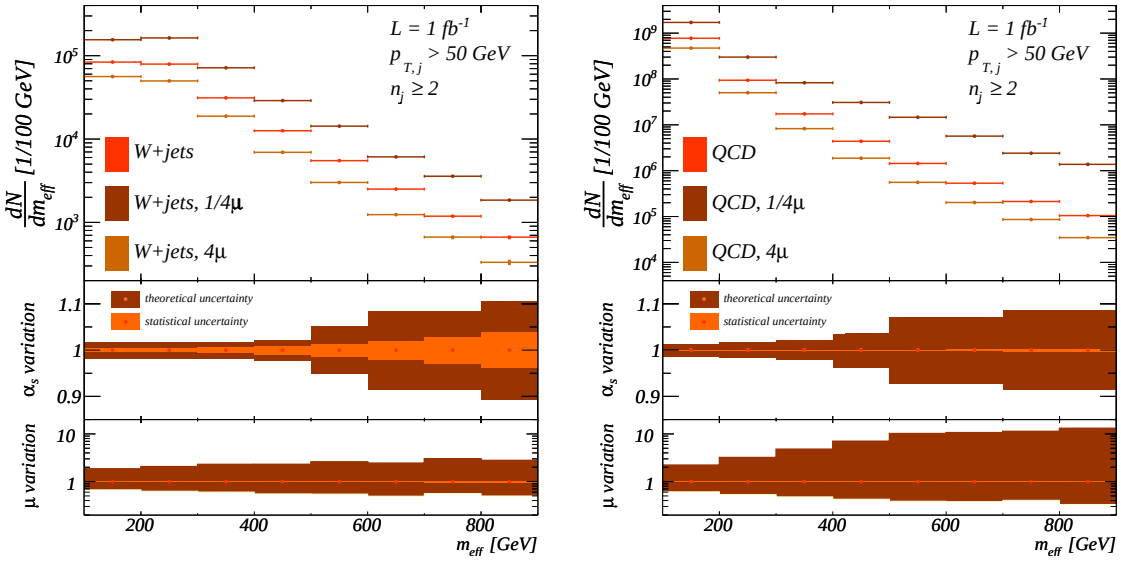

Figure 1.9: Effective mass distribution for W+jets and QCD jets production. Only the jet cuts given in Eq.(1.15) are applied. The second panels show the parametric uncertainty due to a consistent change of $\alpha_s(m_Z)$ between 0.114 and 0.122. The third panels show a consistent scale factor variation which can be experimentally constrained and should not be considered a theory uncertainty.

is the effective mass [67], which for exclusive jet multiplicities we define as

$$m_{\text{eff}} = \not{E}_T + \sum_{\text{all jets}} p_{T,j} \; , \tag{1.20}$$

including all jets fulfilling Eq.(1.15). This definition is neither optimized to take into account a correlation between hard jets and the missing-energy vector nor to remove hard initial-state radiation. Instead, Eq.(1.20) makes a minimal set of assumptions to avoid sculpting the background distribution.

Just like the n_{jets} distribution m_{eff} in the Standard Model cannot be reliably predicted by parton-shower Monte Carlos. Jets entering the sum in Eq.(1.20) we have to understand over their entire transverse-momentum spectrum. CKKW [45,61] or MLM [46] matching is therefore the most adequate approach for simulating m_{eff}.

Exactly following the treatment of the n_{jets} distribution in Section 1.2.2 we estimate two sources of theory uncertainties, the parametric error from varying the strong coupling and the scale-variation systematics. To not be limited by statistics of our background samples we for now discard the missing energy cut and the lepton veto and instead study the fully inclusive processes. In Figure 1.9 we present the m_{eff} distribution for W+jets and the QCD jets production with $n_{\text{jets}} \geq 2$. The same way as in Figure 1.5 we show the relative impact of the two sources of uncertainty in the lower panels. The parametric error from $\alpha_s(m_Z)$ ranges well below 20% even towards large values of m_{eff}. For the electroweak process this is of similar size to the expected statistical error for an integrated luminosity of 1 fb^{-1}. As expected, towards large m_{eff} the error band increases, but not dramatically.

In contrast, the scale-factor variation $\mu/\mu_0 = 1/4 - 4$ has a huge effect on the m_{eff} simulation, essentially rendering it unpredictive. For values above $m_{\text{eff}} = 500$ GeV the error bands become large enough to make it impossible to extract new physics from this observable, were we to consider the scale variation a proper theory error. However, measurements of multi-jet rates and other jet observables at Tevatron and LHC indicate that for the case of SHERPA this scaling factor is approximately one [72]. Measuring the staircase-scaling factors even more precisely with the 2011 LHC data will further constrain the scale ambiguities underlying our QCD simulations – allowing us to make reliable predictions for e.g. the m_{eff} observable.

To see the impact of m_{eff} in searches for supersymmetry we show the m_{eff} distribution for exclusive 2-jet and 3-jet events in Figure 1.10. It includes Standard-Model backgrounds as well as

the supersymmetric signal. All jet-selection and background-rejection cuts specified in Eqs.(1.14) and (1.15) are applied. As mentioned before, QCD jets are the by far dominant channel. Only for $m_{\text{eff}} > 800$ GeV the signal starts overcoming the backgrounds. The statistical uncertainty for 1 fb^{-1} we indicate by the shaded regions in the lower panels. It is worth noticing that the signal+background sample when compared to the pure background sample exhibits a maximum at around $m_{\text{eff}} \sim 1.1$ TeV. This scale corresponds to the squark and gluino masses which for pair production add to 1100 to 1200 GeV. This means that the m_{eff} distribution for exclusive jet multiplicities can serve as background rejection as well as a measure for the mass scale of the new heavy colored states.

1.2.5 Autofocusing

Following our results in the previous sections we should be able to use the shapes of the n_{jets} and m_{eff} distributions to extract a supersymmetric signal from the now quantitatively understood Standard Model backgrounds. Given that the two distributions are affected independently by the color structure of the new physics sector and by its mass scale(s) we will assess the power of the two-dimensional n_{jets} vs m_{eff} correlations in extracting a discovery or an exclusion. Such a two-dimensional shape analysis is the natural second step after the first completely inclusive searches based on counting events. According to Sections 1.2.2-1.2.4 systematic experimental uncertainties will start dominating for luminosities around $\mathcal{O}(1$ fb$^{-1})$. Since those are subject to continuous refinement during data taking and need to be addressed within a full detector simulation study we limit ourselves to statistical uncertainties for a given luminosity. While this means that we will not obtain reliable estimates for the discovery reach, we will see that it allows us to discuss the main benefits and limits of the proposed analysis.

As supersymmetric reference models we choose the benchmark point SPS1a, two variations of it, and SPS4. Again, we only apply the cuts given in Eqs.(1.14) and (1.15) and use the exclusive definition of n_{jets} and m_{eff}. For the m_{eff} distribution we choose a binning of 100 GeV, which approximately reflects the experimental resolution towards large m_{eff}.

For given background and signal+background hypotheses we use a binned log-likelihood ratio to compute statistical significances assuming statistically uncorrelated bins

$$\log Q = \sum_{\text{bins}} \left[n_i \log \left(1 + \frac{s_i}{b_i} \right) - s_i \right] \; . \tag{1.21}$$

It includes the luminosity via the signal and background event numbers s_i and b_i in each bin. While it avoids the limitations of $S/\sqrt{B}$ in regions requiring Poisson statistics it approaches a Gaussian limit for each individual channel when the bin content becomes large. Some features of this well established approach we summarize in Appendix B. Applying a "simple hypothesis test" tells us how likely it is that the background-only hypothesis fakes the predicted signal+background distributions as a statistical fluctuation, i.e. we define the p-value as the SPS1a likelihood ratio's median. The likelihood ratio given in Eq.(1.21) we compute for the exclusive n_{jets}, m_{eff}, and two-dimensional (n_{jets}, m_{eff}) distributions. In this two-dimensional plane the definition of m_{eff},

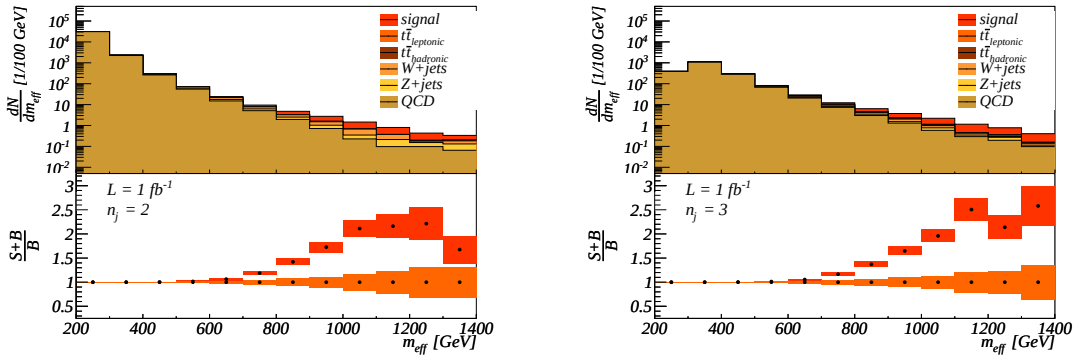

Figure 1.10: Effective mass distribution for exclusive 2-jet and 3-jet events for Standard-Model backgrounds and the supersymmetric signal using the SPS1a parameter point. We assume a center-of-mass energy of 7 TeV with an integrated luminosity of 1 fb^{-1} and apply all cuts in Eqs.(1.14) and (1.15).

	signal significance for 35 pb^{-1}
inclusive	$0.2\,\sigma$
n_{jets} (1D)	$1.6\,\sigma$
m_{eff} (1D)	$3.3\,\sigma$
$(n_{\text{jets}}, m_{\text{eff}})$ (2D)	$4.6\,\sigma$

Table 1.4: Confidence levels for the signal plus background sample ruling out the background-only hypothesis based on one and two dimensional log-likelihood distributions. The supersymmetric mass spectrum is given by SPS1a.

following Eq.(1.20), only includes exactly n_{jets} jets. With this completely exclusive definition of n_{jets} and m_{eff} we ensure that the sum over all bins in the (n_{jets}, m_{eff}) reproduces the total cross section.

Considering this correlation is similar in spirit to the ($\not{E}_T, H_T$) analysis proposed in Ref. [83]. However, first we focus on the n_{jets} and m_{eff} distributions because in Sections 1.2.2-1.2.4 we have shown that we can quantitatively understand the staircase scaling behavior of the Standard Model backgrounds and translate its precision into other variables. In addition, as we will see in this section these two variables play a special role, as they not only distinguish signals from backgrounds, but also contain information on the structure of the underlying new-physics model. As mentioned above, for the sake of a proof of concept we ignore all uncertainties except for statistical experimental errors, to avoid correlations in the definition of the log-likelihood.

We can expect from Figures 1.8 and 1.10 that the rate in each individual n_{jets} bin is dominated by Standard-Model processes at low m_{eff}. Most likely, this region will be the control region to normalize the QCD and W/Z+jets backgrounds. With the exception of hadronically decaying top pairs all Standard-Model channels will then show a simple decrease in both directions of the two-dimensional (n_{jets}, m_{eff}) plane which we can predict following the arguments in Sections 1.2.2-1.2.4. The signal contribution will become visible only once m_{eff} reaches the mass range of the particles produced.

In Table 1.4 we compare the statistical significances for the supersymmetric SPS1a parameter point at 7 TeV center-of-mass energy for the various analysis strategies: first, we show the results based on the total production rates after the inclusive cuts of Eqs.(1.14) and (1.15). As expected, including the signal events leaves us completely consistent with the background-only hypothesis. Next, the likelihood ratio computed from the n_{jets} distribution gives rise to sizable deviations from the background for integrated luminosities as small as 35 pb^{-1}. The one-dimensional m_{eff} distribution turns out to be an even better discriminator. It gives us more than twice the n_{jets} significance, namely $3.3\,\sigma$ for $\mathcal{L} = 35$ pb^{-1}. The highest significant discriminative power we obtain for the two-dimensional binned (n_{jets}, m_{eff}) case. This is a direct consequence of the additive binned log-likelihood given in Eq.(1.21).

Beyond the relevance of the (n_{jets}, m_{eff}) distributions to extract new particles from backgrounds, we can utilize it to study signal properties. Above, we argue that new physics contributions to n_{jets} will only appear once m_{eff} reaches the mass scale of the sum of both heavy particles produced. However, this only happens if the exclusive n_{jets} value allows us to include the decay

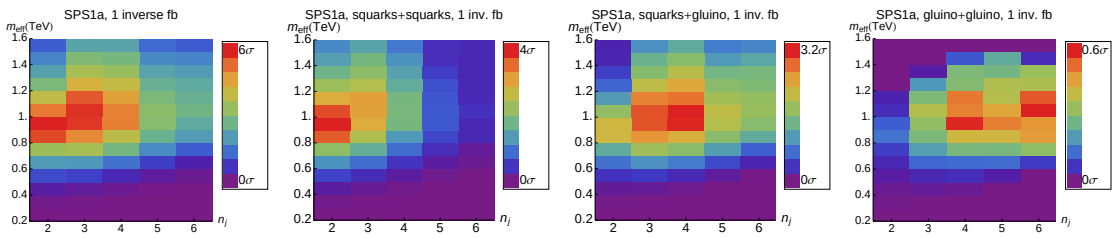

Figure 1.11: Log-likelihood contributions over the (n_{jets}, m_{eff}) plane for the supersymmetric signal using the SPS1a spectrum. The color code is normalized to different maximum significances.

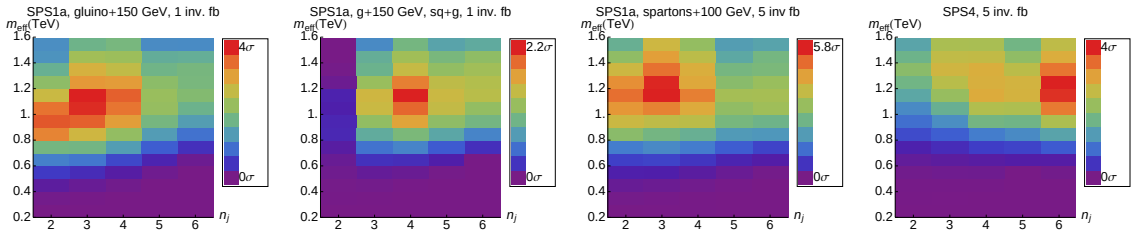

Figure 1.12: Log-likelihood distributions over the ($n_{\rm jets}$, $m_{\rm eff}$) plane for the different supersymmetric spectra. Due to the smaller signal rates we present results for an integrated luminosity of 5 fb^{-1}.

jets contributing to $m_{\rm eff}$. Hence, the new physics contributions to the two observables will show a correlation based on the mass and decay channels of the new particles produced. The decay channels can typically be linked to the color charge of the new particles if we assume that the missing energy particle cannot carry color charge. Color triplets will tend to decay to one hard quark jet while color octets with their diagonal coupling to gluons will radiate two quark jets. This means breaking down the binned log-likelihood ratio over the fully exclusive ($n_{\rm jets}$, $m_{\rm eff}$) plane and keeping track of the individual contribution of each bin will automatically focus our search on the appropriate properties of the particles we are looking for.

This statement is not limited to supersymmetry, the SPS1a parameter point or any other assumption about the signal. It can be applied to general physics beyond the Standard Model with strongly interacting new particles and a stable dark matter candidate. In Figure 1.11 we show the contributions of the individual bins to the summed log-likelihood ratio for all signal events combined and split into three production processes. The maximum significance automatically reflects SPS1a's decay paradigm $\tilde{q}\tilde{q}^{(*)} \to 2$ jets and $\tilde{q}^{(*)}\tilde{g} \to 3$ jets, and $\tilde{g}\tilde{g} \to 4$ jets, know already from Figure 1.7. The first two channels we can study using an integrated luminosity of 1 fb^{-1}. Squark pair production is dominant because at the LHC it includes a quark-quark initial state. Associated production, which often is the dominant channel at the LHC, has a comparable statistical yield and features a slightly higher $m_{\rm eff}$ range. Both channels combined define the diagonal correlation we see for the combined signal events.

Gluino pair production has the smallest production rate and therefore becomes subleading in the combined supersymmetry sample. However, for this channel we can best follow the imprint of higher jet multiplicities. Due to its large mass and its color charge gluino pairs produce significantly more jet radiation which we can resolve for a sufficiently low $p_{T,\rm min}$ threshold. For $p_{T,\rm min} = 50$ GeV we might just capture the first decay jet from the gluino cascade, reflecting the mass hierarchy $m_{\tilde{g}} - m_{\tilde{q}} \sim 60$ GeV. The peak in the log-likelihood plane around $n_{\rm jets} = 4$ results from the maximum in the $\tilde{g}\tilde{g}$ production cross section. For $n_{\rm jets} = 5$ the background is still large compared to the signal, but dropping at an exponential rate it gets surpassed for $n_{\rm jets} = 6$, explaining the structure we observe in Figure 1.11.

Finally, we can study how changes to the new physics spectrum are reflected in the significances computed from the binned log-likelihood $Q(\, n_{\rm jets},\, m_{\rm eff})$. We investigate three different supersymmetric mass spectra : first, we increase only the gluino mass by 150 GeV with respect to SPS1a ($\sigma^{\rm NLO}_{\rm SUSY} = 2.69$ pb according to Prospino2.1 [78]); second, we increase all colored-sparticle masses by 100 GeV with respect to SPS1a ($\sigma^{\rm NLO}_{\rm SUSY} = 1.63$ pb); third, we consider the SPS4 benchmark [82] with an inverted mass hierarchy $m_{\tilde{q}} \sim 750$ GeV $> m_{\tilde{g}} \sim 730$ GeV ($\sigma^{\rm NLO}_{\rm SUSY} = 0.83$ pb). All of these cross sections are significantly smaller than for SPS1a with its $\sigma^{\rm NLO}_{\rm SUSY} = 4.68$ pb, which means we increase our nominal luminosity to $\mathcal{L} = 5$ fb^{-1}.

In Figure 1.12 we clearly see the effect of the increased gluino mass. The $m_{\rm eff}$ peak for associated squark-gluino production moves to larger values, as does the $n_{\rm jets}$ maximum. However, because the balance between squark pair production and associated squark-gluino production shifts into the direction of the squark pairs, this effect is not quite as pronounced. The second scenario with increased squark and gluino masses leads to a pronounced maximum at larger $m_{\rm eff}$. Due to the smaller signal cross section the sensitivity in particular in the $n_{\rm jets} = 2$ bins gets considerably diminished, appearing as a shift towards higher $n_{\rm jets}$ values. For the all-hadronic search in the SPS4 parameter point longer decay chains for gluinos through bottom squarks appear in the high

n_{jets} bins only.

The SPS4 case illustrates that $n_{\text{jets}} = 6$ does not have to be the maximum jet multiplicity we need to consider. Once we rely on a combination of data and Monte Carlo methods to describe the n_{jets} staircase scaling for background processes we can extend our analyses to very large jet multiplicities. On the other hand, Figure 1.11 also clearly indicates that for example in the SPS1a parameter point the optimal signal extraction strategy by no means requires us to go to very large jet multiplicities. For the SPS1a parameter point the two-jet bins are leading contributions to the total significance.

Outlook

Multi-jet events with and without a weak gauge boson are the dominant backgrounds to inclusive searches for example for supersymmetry. Simulating them with traditional parton shower Monte Carlos does not lead to reliable results. This changes when we make use of jet merging to predict the shapes of multi-jet backgrounds.

One of the most challenging distributions to describe theoretically is the inclusive or exclusive number of jets per event. From data we know that the inclusive W/Z+jets distributions without rigid cuts follow a *staircase scaling* with constant jet ratios $R = \hat{\sigma}_{n+1}/\hat{\sigma}_n$. We have shown that staircase scaling in the inclusive jet rates is equivalent to the same scaling in exclusive rates. Nowadays, the exclusive scaling we can compare to the predictions of jet merging Monte Carlos, like SHERPA. Moreover, we have seen that QCD jet production shows an even more pronounced scaling than W/Z+jets production to jet multiplicities of $n_{\text{jets}} = 8$.

We studied the validity of the staircase scaling behavior, the theory uncertainties in the n_{jets} distributions, its link to other multi-jet observables, and the application of jet-exclusive observables to new physics searches and found:

1. While we cannot derive the staircase scaling of the n_{jets} distribution from first principles we can reproduce it using the appropriate Monte Carlo tools. This includes the scaling feature itself, a careful error analysis, and the scaling violation effects towards large values of n_{jets} due to phase space restrictions.

2. The theory uncertainty on the staircase scaling consists of tunable parameters like scale factors in the factorization and renormalization scales and on parametric errors like the dependence on α_s. The latter are small. The scale factor hugely overestimates the error and should be thought of as a tuning parameter for the different jet merging implementations. For SHERPA it comes out close to unity.

3. The scaling parameter $R = \sigma_{n+1}/\sigma_n$ depends on the hard process and on kinematic cuts. Both effects we can reliably predict using Monte Carlos, as we have shown for the W/Z+jets and pure jets cases. The γ+jets case we postpone to a later study with more specific details for the LHC experiments [84].

4. These simulations of the staircase scaling in multi-jet processes can be easily combined with data driven techniques, giving us the over-all normalization and a cross check for the first n_{jets} bins. Statistically limited regions of phase space will become accessible via simulations, including a reliable error estimate.

5. Understanding staircase scaling of multi-jet processes allows us to predict other multi-jet variables, like the effective mass m_{eff}. Again, this includes a proper treatment of theory uncertainties. In addition, the completely inclusive definition of m_{eff} removes dangerous artifacts due to the usual truncations.

6. Based on for example the n_{jets} vs m_{eff} correlation for a fixed $p_{T,\text{min}}$ we can define a likelihood-based analysis avoiding model or spectrum specific background rejection cuts. Such shape analyses in multi-jet search channels are the natural extension of the early inclusive ATLAS and CMS searches.

Of course this simple first approximation to the exclusive zero-lepton search for jets plus missing energy is not the only application of such methods. Searches including leptons, b tags, or hard photons will benefit from the same treatment, as long as they include non-negligible numbers of jets. The same is true for hadronically decaying top quarks in the Standard Model.

1.3 Scaling features in QCD

In Sec. 1.2 we study staircase scaling and use it to estimate the standard model backgrounds in new physics searches. However, one very important ingredient is missing: why do we observe staircase scaling? If we are to use this pattern for data driven background estimation and new physics searches, we need to fully understand its emergence as well as its limitations. For this theoretical enterprise it is necessary to know that there exits a second pattern, too: Poisson scaling. It is defined by

$$R_{\frac{n+1}{n}} = \frac{\bar{n}}{n+1} = \frac{\sigma_{n+1}}{\sigma_n}. \tag{1.22}$$

Radiation featuring this pattern is following a simple Poisson distribution. A fact well known from QED [22].

1.3.1 Data

Before we move on to explain scaling patterns from theory, let us shortly collect the current status of the data and in which phase space regions they are observed. The ATLAS collaboration published a study in Z plus jets [123] and compared their data with the staircase scaling as well as the Poisson scaling [22, 24] hypothesis. We show their results in Fig. 1.13. We see that both features are realized. The Poisson pattern, however, turns into the staircase pattern for high multiplicities. The kinematic regions defined by the experimental analysis cuts are called democratic selection for the staircase case and large logarithm for the Poisson case. We call the former democratic, because there is only one scale involved: the $p_T^{\min}$ defining the jets. The other scenario is called large logarithm, because we ask for a high p_T leading jet. This introduces a second scale and leads to the occurrence of large logarithms as explained in Sec. 1.1.

We analyze these two phase space regions using the generating functional formalism introduced in Sec. 1.1. We follow a systematic approach. After defining both scaling patterns formally we study different possible matrix element configurations for the two jet emission case in the parton shower approximation. This gives us a first hint why each pattern is associated with a certain phase space regime. Next we study final state radiation in an e^+e^- collider environment. We find analytic solutions for the generating function producing staircase and Poisson scaling. The translation to hadron colliders then contains two steps. We first translate our findings from (E, ϕ, θ) coordinates to the corresponding hadron collider coordinates (p_T, ϕ, η). In a second step we add the effects of initial state partons, the pdfs.

The following subsections "Multi-jet rates" and "Final-state parton shower" have already been published in [29] together with Erik Gerwick, Steffen Schumann and Tilman Plehn.

1.3.2 Multi-jet rates

Multi-jets final states are ubiquitous at hadron colliders. QCD jet radiation mostly off the initial state partons has huge impact on almost every LHC analysis. For example in Higgs searches an accurate description of the jet recoil against the Higgs boson allows for an efficient rejection of many backgrounds [89]. In top pair production and single top production the identification of QCD recoil jets on the one hand and decay jets on the other hand is one of the limiting factors in precision analyses [90, 91]. New physics searches largely rely on hard decay jets of new strongly interacting particles, which makes them vulnerable in the case of soft decay jets hidden in the QCD jet activity [53, 92, 93]. Understanding the jet multiplicity and the jet spectra from QCD radiation is a core ingredient to improving any of these analyses.

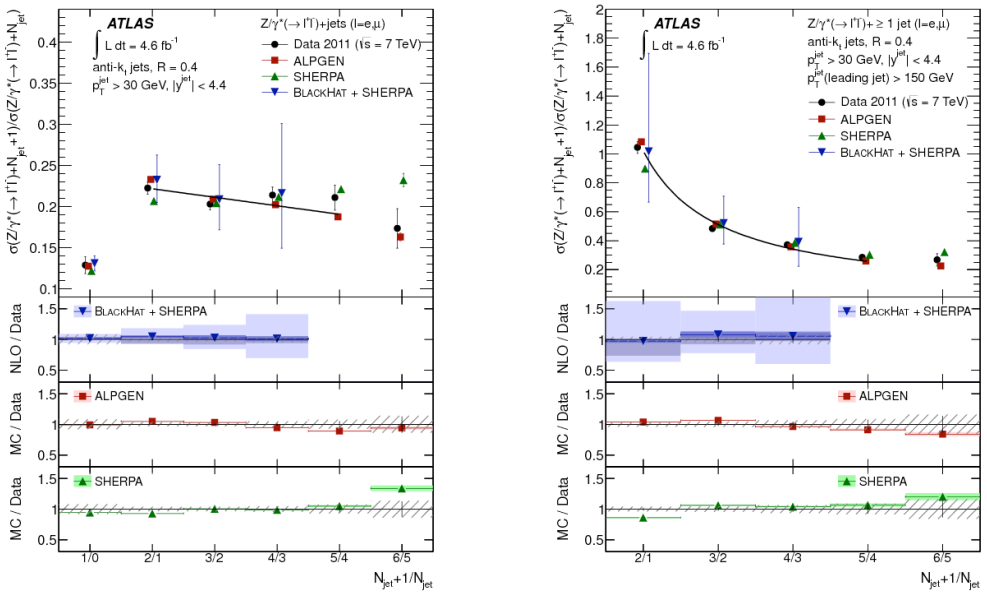

Figure 1.13: ATLAS data for the Z plus jets channel. Plotted are the jet cross section ratios. Left: staircase scaling. Right: Poisson scaling. Figure taken from [123].

Even though jet radiation seems to follow simple patterns [95, 97], theoretical predictions for multi-jet observables in perturbative QCD are challenging. It is clear that the radiation of relatively hard quarks and gluons is a direct consequence of the hierarchy between the large proton-proton collider energy and the typical electroweak scale of the partonic interaction [98, 99]. Numerically, we can combine the QCD parton shower with hard matrix element calculations, to predict jet radiation patterns over wide phase space regions [45, 46, 100]. Analytically, Sudakov factors and generating functionals can be used to describe QCD jet radiation [101]. A careful comparison of analytical and numerical approaches to LHC data would allow us to determine the strengths and limitations of the underlying theoretical concepts: fixed-order perturbation theory, parton showers, and resummation based on generating functionals.

In fixed order perturbation theory leading-order jet rates are available for effectively arbitrarily high multiplicities [48,102]. The number of jets we can consider is limited only by computing power. However, leading order predictions suffer significant shortcomings when it comes to precision. The renormalization scale dependence as a measure for the theoretical uncertainty grows with each power of α_s when we add a final state jet. High powers of the logarithmic scale dependence $(\log \mu_R)^n$ mean that such cross section predictions can only be considered an order-of-magnitude estimate. In addition, in the presence of phase space constraints large logarithms spoil the convergence of fixed order perturbation theory. For example jet vetoes will induce sizable logarithms of ratios of leading jet p_T or masses (m_H or m_Z) to additional resolved jets (experimentally relevant down to $20 - 30$ GeV) [103, 104].

Next-to-leading order computations ameliorate the scale dependence and capture an additional logarithm. However, we are still limited in the available final-state multiplicity, e.g. pure jets are available for $n_{\text{jets}} \leq 4$ [105], $n_{\text{jets}} \leq 4$ in association with W/Z bosons [106,107], for $t\bar{t}$ production and Higgs production in gluon fusion NLO corrections are known for $n_{\text{jets}} \leq 2$ [108,109]. However, over the past few years this field has progressed enormously. As a consequence for Standard Model processes a similar level of automation as for leading-order calculations is within reach [110]. An approach applicable for general New Physics extensions, though limited to $2 \to 2$ processes, has been presented in [111].

Although NLO calculations contain one additional power of enhanced logarithms, this might not be sufficient for high jet multiplicities. At NNLO, although there has been an enormous amount of recent development [112], the number of fully differential calculations is limited, and an automated implementation is not foreseeable in the near future.

On the other hand, we know that jet radiation is enhanced by traceable logarithms. This makes improved predictions for QCD observables based on resummation possible. The general strategy

is to redefine the perturbative series from powers of α_s to including the relevant logarithms; the simplified structure of these enhanced terms then allows for a resummation to all orders. Once the resummed form is known we can match onto a fixed order calculation and avoid double-counting. For Sudakov-type logarithms a general method for this type of resummation is available [113], and for particular event shape observables an automated approach exists [114]. In LHC analyses, the resummation of finite logarithms in the presence of a jet-veto scale is of interest [103, 115].

A numerical approach to resummation is provided by parton-shower simulations [116]. It is automated in the multi-purpose Monte Carlo generators Pythia [50], Herwig [51] and Sherpa [47] to leading order in the strong coupling combined with the resummation of leading collinear logarithms (LO/LL). This method differs from the previous approaches in that the full spectrum of final state partons or hadrons is produced explicitly. While the parton shower is well defined for relatively small transverse momenta of the jets it is not applicable for hard jet radiation. However, this limitation is overcome by the CKKW [45], MLM [46], and CKKW-L [117] jet-merging algorithms, that incorporate the tree-level matrix-element corrections for the first few hardest emissions [116, 118].

A complementary strategy is provided by the MC@NLO [119] and POWHEG [120] approaches, that realize the matching of NLO calculations with parton showers. While these methods guarantee NLO/LL accuracy only the first/hardest shower emission gets corrected by the real-emission matrix element. Higher jet multiplicities are described in the parton-shower approximation only. First attempts to combine the NLO/LL approaches with the tree-level merging ansatz have been reported recently [121]. An unprecedented level of sophistication for predicting multi-jet final states is achieved by the promotion of merging algorithms to next-to-leading order accuracy [122].

Even though we can nowadays simulate multi-jet events, a detailed understanding of inclusive or exclusive n_{jets} distributions at the LHC is still missing. Its universal features have been studied since 1985 [95]. Scaling patterns can be conveniently displayed in the ratio of successive exclusive jet cross-sections

$$R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} = \frac{P_{n+1}}{P_n} \qquad \text{with} \qquad P_n = \frac{\sigma_n}{\sigma_{\text{tot}}} \,. \tag{1.23}$$

We define the jet multiplicity n as the number of jets *in addition* to the hard process, e.g. σ_1 for pure QCD di-jets is experimentally a 3-jet final state. Jets which are part of the hard process are not included in the scaling analysis because they do not arise from single QCD emissions.

Two patterns provide limiting cases for most LHC processes and are referred to as staircase and Poisson scaling. Staircase scaling is defined as constant ratios between the successive multiplicity cross sections

$$R_{(n+1)/n} = R \equiv e^{-b} \,, \tag{1.24}$$

where R and b are constant. The exclusive n-jet rate for this distribution is $\sigma_n = \sigma_0 e^{-bn}$ where σ_0 is the 0-jet exclusive cross section. Staircase scaling for exclusive and inclusive jet rates is equivalent, with identical values of R. For a Poisson distribution with expectation value $\bar{n}$ the rates are

$$P_n = \frac{\bar{n}^n e^{-\bar{n}}}{n!} \qquad \text{or} \qquad R_{(n+1)/n} = \frac{\bar{n}}{n+1} \,. \tag{1.25}$$

More properties of the distributions are described in Ref. [89].

1.3.3 Final-state parton shower

The simulation of a LO/LL event in the parton shower approximation starts with the generation of a single phase space point for the partonic core process. The process' external (colored) lines then act as seeds for the subsequent parton shower evolution. Driven by unitarity they start at the hard process scale μ_F and finish at the shower cutoff scale $\mu_0 \sim 1$ GeV. Hard matrix element corrections for an arbitrary number of additional particles can be added to the parton shower using the above-mentioned matching schemes [45, 46, 117]. However, in this section we treat all additional emissions as coming from the parton shower and disregard matrix element corrections.

The basis of the LL parton shower is the fully factorized form of the collinear matrix element and phase space $d\sigma_{n+1} \sim d\sigma_n \; P_{1\to 2} \; d\Phi_1$. Using this simplification the parton shower remains local, but loses information on spin, color correlations and interference effects in addition to higher order terms neglected in the $1 \to 2$ splitting kernels $P_{1\to 2}$. Besides practicality, one of the benefits

Figure 1.14: Simplest primary (left) and secondary contributions (right) assuming a core process with a hard quark line.

of collinear factorization is that the resummation of LL and some NLL contributions follow very naturally. To see this, we represent the evolution along an individual line by integrating the Sudakov factor over the appropriate virtuality scales from the lower cutoff μ_0^2 to a free hard scale t

$$\Delta_j(t) = \exp\left[-\int_{\mu_0^2}^{t} dt'\, \Gamma_j(t,t')\right] . \tag{1.26}$$

For the regularized splitting kernels Γ_j we use the next-to-leading logarithmic approximation

$$\Gamma_j(t,t') \;=\; c_j\, \frac{\alpha_s(t')}{2\pi t'} \left(\log\frac{t}{t'} - A_j\right) , \tag{1.27}$$

with color factors $c_j = C_F$ (C_A) and the constant terms $A_j = 3/2$ (11/6) for gluon emission off a quark (gluon). The lower cutoff scale μ_0 is omitted in the argument of the Sudakov form-factor. Expanding the exponential we see that Eq. 1.26 represents an arbitrary number of soft and collinearly enhanced emissions, either resolved or unresolved.

To describe a parton-shower simulated event we note that the QCD evolution proceeds as an integration of the product Sudakov along the virtuality t,

$$\mathbf{\Delta}(t) = \prod_{\text{ext lines}} \Delta_j(t) \equiv e^{-\mathbf{\Gamma}} . \tag{1.28}$$

The product defining $\mathbf{\Gamma}$ is over the appropriate factors for each external line, where j denotes the particle flavor. Limiting ourselves to final state splittings this expression only contains evolution kernels as shown in Eq. 1.26, and it is by construction guaranteed to exponentiate with an appropriate expression $\mathbf{\Gamma}$. As long as $\mathbf{\Gamma}$ is fully local and does not depend on previous emissions it is guaranteed to produce a Poisson distribution for the multiplicities. The exponentiated form in Eq. 1.28 immediately identifies $\bar{n} = \mathbf{\Gamma}$. This statement does not depend on the form of $\mathbf{\Gamma}$ or its dependence on the hard scale t. All that matters is that each splitting does not change the subsequent evolution. In the remainder of this paper we define all emissions directly contained in the expansion of Eq. 1.28 as *primary* with respect to the core process.

The first splitting in the parton shower picture defines the single emission probability. Following 1.14 a second emission can then appear from the original leg or off the first emission. For the former, this emission is contained in Eq. 1.28 and does not change the Poisson pattern. The latter changes the exponential; we refer to it as *secondary* with respect to the original hard process. From a scaling perspective the relevant questions are first, what is the relative size of the two contributions; and second if we can change the individual strengths of primary and secondary emissions through kinematic cuts.

In the parton shower approximation we can associate specific integrals over virtuality with individual partonic structures appearing in the final state evolution. An alternative evolution ordered in a consistent variable (e.g. angle) is logarithmically equivalent. Using this formalism the primary contribution to two gluon emission off a hard quark shown in 1.14 is

$$\sigma^{\text{primary}}(\mu_F^2,\mu_0^2) = c^{\text{primary}} \int_{\mu_0^2}^{\mu_F^2} dt\, \Gamma(\mu_F^2,t)\Delta_g(t) \int_{\mu_0^2}^{\mu_F^2} dt'\, \Gamma(\mu_F^2,t')\Delta_g(t') . \tag{1.29}$$

The coefficient c^{primary} which includes the Sudakovs associated with the hard line is process dependent, as this hard line can be either a quark or a gluon. The two external scales are the scale μ_F of the hard process and the lower cutoff scale μ_0. If the primary emissions are strongly ordered

in the evolution variable, the corresponding phase space factor 1/2 is absorbed in c^{primary}. The simplest secondary contribution also shown in 1.14 is,

$$\sigma^{\text{secondary}}(\mu_F^2,\mu_0^2) = c^{\text{secondary}} \int_{\mu_0^2}^{\mu_F^2} dt\,\Gamma(\mu_F^2,t)\Delta_g(t) \int_{\mu_0^2}^{t} dt'\,\Gamma(t,t')\Delta_g(t') \;. \tag{1.30}$$

The splitting kernels in the two expressions only differ in the integral boundaries for the second emission. In the leading logarithmic approximation (in the exponent) for the Sudakov factors, we can perform the integrals in Eq. 1.29 and Eq. 1.30 in terms of error functions. The full expressions are not particularly enlightening, but two specific limits contain crucial information.

(1) $\frac{\alpha_s}{\pi}\log^2\frac{\mu_F}{\mu_0} \gg 1$

In this limit we expand Eq. 1.29 and Eq. 1.30 around $\mu_0/\mu_F \to 0$ and find the leading terms

$$\begin{aligned}
\sigma^{\text{primary}} &= \frac{c^{\text{primary}}}{4}\left[\frac{\alpha_s}{C_A}\log^2\frac{\mu_F}{\mu_0} \;-\; \sqrt{\frac{4\alpha_s}{C_A^3}}\log\frac{\mu_F}{\mu_0} \;+\; \mathcal{O}\left(\frac{\mu_0^2}{\mu_F^2}\right)\right] \\
\sigma^{\text{secondary}} &= \frac{c^{\text{secondary}}}{4}\left[(\sqrt{2}-1)\sqrt{\frac{\alpha_s}{C_A^3}}\log\frac{\mu_F}{\mu_0} \;+\; \mathcal{O}\left(\frac{\mu_0^2}{\mu_F^2}\right)\right] \;.
\end{aligned} \tag{1.31}$$

Their ratio scales like $\sigma^{\text{primary}}/\sigma^{\text{secondary}} \propto \sqrt{\alpha_s}\log\mu_F/\mu_0$, i.e. the primary emissions are logarithmically enhanced. In the limit of a large logarithm (high single emission probability) the distribution of final state emissions are increasingly primary, and therefore give a Poisson distribution.

Physically interpreting Eq. 1.31, a second logarithm in the secondary contribution would come from the right-most Sudakov of Eq. 1.30. However, it has vanishing support for $\mu_F \to \infty$ and does not appear in the approximate result. The emitted gluon in this case spans a vanishing relative fraction in virtuality space where it may emit an additional parton.

(2) $\frac{\alpha_s}{\pi}\log^2\frac{\mu_F}{\mu_0} \ll 1$

Taking this limit of Eq. 1.29 and Eq. 1.30, we find

$$\begin{aligned}
\sigma^{\text{primary}}(\mu_F^2,\mu_0^2) &= c^{\text{primary}}\frac{\alpha_s^2}{4(2\pi)^2}\log^4\frac{\mu_F}{\mu_0} \;+\; \mathcal{O}\left(\alpha_s^3\log^6\frac{\mu_F}{\mu_0}\right) \\
&= 6\,\frac{c^{\text{primary}}}{c^{\text{secondary}}}\,\sigma^{\text{secondary}}(\mu_F^2,\mu_0^2) \;.
\end{aligned} \tag{1.32}$$

The two contributions become logarithmically equivalent and differ by an $\mathcal{O}(1)$ constant depending primarily on color factors. In this regime the emission probability is small and the final state is selected democratically. The formerly Poisson scaling pattern receives large contributions from subsequent or secondary splittings. Note that to justify the logarithmic expansion we still require $\log^2\mu_F/\mu_0 > 1$ but not large enough to spoil the small emission probability.

1.3.4 Scaling patterns at lepton colliders

Let us now explore these two limits via the generating functional formalism. We consider a lepton collider environment, which means that we only study final state radiation. The generalized k_T algorithms at lepton colliders are defined with the following distance measure between two particles i and j [38]:

$$d_{ij} = \min\left(E_i^{2k}, E_i^{2k}\right)\frac{1-\cos\theta_{ij}}{1-\cos\theta_R}. \tag{1.33}$$

The specific jet algorithm we use is defined by k, where $k=1$ corresponds to the k_T algorithm, $k=0$ is the C/A algorithm, and $k=-1$ corresponds to the anti-k_T algorithm. A jet algorithm than works as follows:

1. find $d^{\min} = \min_{ij\in\text{particles}}\left(d_{ij}, E_i^{2k}\right)$

2. if $d^{\min}$ is of the type E_i^{2k}: call i a jet and remove it from the list of particles, go back to (1.)
3. if $d^{\min}$ is of the type d_{ij}: replace i and j by $i+j$, go back to (1.)
4. iterate until all particles are called jets.

As in the hadron collider case we need a lower cut off. Therefor, we demand that only jets with an energy above some E_R are taken into account. It is convenient to introduce dimensionless parameters $e = E/E_R$ and $\xi_{(R)} = 1 - \cos\theta_{(R)}$. In these parameters the DGLAP equation for the generating functional reads [124]

$$\Phi_i(e,\xi) = u\Delta_i(e,\xi) + \sum_{i\to jk} \int_{\xi_R}^{\xi} \frac{d\xi'}{\xi'} \frac{\Delta_i(e,\xi)}{\Delta_i(e,\xi')} \int_e^1 \frac{\alpha_s(k_T^2)}{2\pi} P_{i\to jk}(z)\Phi_k(e,\xi')\Phi_j(ze,\xi'). \tag{1.34}$$

The interesting point to note is that this prescription separates the soft and collinear divergences. In the original paper [29] we use the notion of scales as in the paragraphs before. However, the generating functional is now connected to the actual jet algorithms used in LHC phenomenology. In addition, it turns out that this description is much more powerful and we gain additional knowledge. Remember, that in the exact collinear limit all evolution variables are equivalent. To study scaling features, especially in the high multiplicity limit, we need the exponentiated form (derived in app. C.2) of the evolution equation. It is

$$\Phi_i(e,\xi) = u\,\exp\left(\int_{\xi_R}^{\xi} \frac{d\xi'}{\xi'} \int_{1/e}^1 dz \frac{\alpha_s(z,\xi')}{2\pi} \sum_{i\to jk} P_{i\to jl}(z)\left[\frac{\Phi_j(e,\xi')\Phi_k(\mathcal{E}(z),\xi')}{\Phi_i(e,\xi')} - 1\right]\right), \tag{1.35}$$

where $i,j,l \in \{q,\bar{q},g\}$. The function $\mathcal{E}(z) = ze$ except for $g\to q\bar{q}$ where it is just e. The form is very similar to the one found for the so called Durham algorithm [24,25][38], which we used in [29]. From this result we can compute resummed jet rates to all orders. As we are mostly interested in resumming logarithms we can use the logarithmic expansion of the running coupling [124].

$$\frac{\alpha_S(z,\xi')}{\pi} = \bar{\alpha} - b_0\bar{\alpha}^2\left(2\log z + \log\frac{\xi'}{\xi}\right). \tag{1.36}$$

where $\bar{\alpha} = \alpha_s(1,\xi)/\pi$ is defined in terms of the coupling at the hard scale. Last but not least, note that the Sudakov form factor in these evolution variables reads

$$\Delta_i(e,\xi) = \exp\left(-\int_{\xi_R}^{\xi} \frac{d\xi'}{\xi'} \int_{1/e}^1 dz \frac{\alpha_s(z,\xi')}{2\pi} \sum_{j,l} P_{i\to jl}(z)\right). \tag{1.37}$$

Let us now expand Eq. (1.35) in the two phase space regimes we are interested in.

Poisson limit

We know that if we induce large logs we expect Poisson scaling. The generalized k_T version depends in fact on two scale choices. The spacial jet resolution $\xi_R = 1-\cos R$ and the allowed energy range defined by $E_R = p_T^{\min}$. In the limit $1 \geq \xi \gg \xi_R$ the whole integral is dominated by the $\xi' \approx \xi_R$ part. From the last paragraph about primary and secondary emissions, we also know that in this limit we can approximate the evolution of the emitted parton with $\Phi_i = u$. The limit $1/e \to 0$ is a little bit more complicated because it depends on the structure of the splitting kernels. These have poles of the form $1/z$ which means that the $z \approx 0$ region contributes most. Therefor, we have $ze \to 1$. In both cases we find

$$\Phi_i(e,\xi) = u\,\exp\left[[u-1]\int_{\xi_R}^{\xi} \frac{d\xi'}{\xi'} \int_{1/e}^1 dz \frac{\alpha_s(z,\xi')}{2\pi} \sum_{j,l} P_{i\to jl}(z)\right]. \tag{1.38}$$

[38] That is an older version of the k_T algorithm, where the soft and collinear structure are not separated. It is used to compute jet rates for the Opal experiment and does agree well with data [25].

This equation produces Poisson scaling, see App. C.1. However the $g \to q\bar{q}$ splitting drops out for the energy limit case. Its splitting kernel is not divergent but goes to zero as $z \to 0$.

Implicitly we have truncated the generating functional at some jet multiplicity. This allows us to ignore the pieces, which are not in the dominant region we use to simplify the evolution equation. After a certain amount of additional radiation we have moved down the scale and our assumption of far apart integral boundaries fails. At this point also the Poisson scaling has to break down. Therefor, Poisson scaling patterns are only expected for low multiplicities.

Staircase scaling

Staircase scaling occurs for democratic scale configurations. In the data example given in Fig. 1.13 we have a fixed jet radius, corresponding to fixed ξ_R. The only other scale is $p_T^{\min}$. Therefor we study Eq. (1.35) in the $e \to 1$ limit. This is exactly the democratic limit. We also constrain our self to the pure Yang-Mills case. The results for starting with a quark can be restored easily. In accuracy this means that our findings are LL results. Note that we can write eq. (1.35) in the form

$$\Phi_g(e,\xi) = u \exp\left(\int_{\xi_R}^{\xi} \frac{d\xi'}{\xi'} \int_{1/e}^{1} dz \frac{\alpha_s(z,\xi')}{2\pi} P_{g\to gg}(z) \left[\Phi_g(e,\xi') + \sum_{l=1}^{\infty} \frac{(e(z-1))^l}{l!} \frac{d^l \Phi_g(e,\xi')}{de^l} - 1\right]\right). \tag{1.39}$$

We Taylor expand around $z_0 \approx 1$ but with $z < 1$. Now, taking only the $l = 1$ term into account, we are able to find a closed solution for Φ_g (see app. C.3), which is

$$\begin{aligned}\Phi_g(e,\xi) &= \frac{1}{1 + \frac{(1-u)}{u\Delta_g(e,\xi)} - (u-1)\chi(e,\xi)} \\ &\approx \frac{1}{1 + \frac{(1-u)}{u\Delta_g(e,\xi)}} + \frac{(u-1)\chi(e,\xi)}{(1 + \frac{(1-u)}{u\Delta_g(e,\xi)})^2}.\end{aligned} \tag{1.40}$$

By taking successive u derivatives we find that the jet rate ratios follow

$$R_{\frac{n+1}{n}} = (1 - \Delta_g(e,\xi)) \left[1 + \underbrace{\frac{1}{\frac{(1-\Delta_g(e,\xi))^3}{\chi(e,\xi)\Delta_g^2(e,\xi)} - \frac{1}{\Delta_g(e,\xi)} - 2}}_{\mathcal{B}(e,\xi)} + (n+1)\right]. \tag{1.41}$$

This is a very interesting result. First of all we find that staircase scaling is predicted by theory. It is not an accident. However, it only occurs for high multiplicities. For lower multiplicities we find staircase scaling breaking terms, which fall off like $1/n$. This is exactly the behavior we expect from the low multiplicities following a Poisson pattern. However, the $1/n$ behavior is modified by the quantity $\mathcal{B}$. It is therefor not the Poisson scaling part of the distribution, but an approximation to the intermediate part, where the Poisson turns into a staircase pattern.

Phase space effect

It is obvious that at some point we cannot produce more jets, because there is just no space in the detector left. Note that at LL the generating functionals for all jet algorithms are equivalent [124]. In addition the anti-k_T algorithm has the beautiful property, that its jets are circular with radius R [38]. However, phase space will not just be covered by circles of size R. Jets, if viewed as circles of size R in the anti-k_T algorithm, will overlap. Therefor, there is more space for jets then naively given by $4\pi/\pi R^2$. From the divergent structure of QCD we expect an additional jet 1 to occur exactly at the border of another harder jet 2. The area for such a configuration is calculable and we find $A_1/A_2 = 0.4$. Taking this information into account, phase space gets covered with an effective jet area, which is smaller then the full possible area. This effective area, however, depends on the

ratio between full area jets and partial area jets. In the anti-k_T algorithm jets are built by a hard parton seed, which sucks up all the soft stuff around it. The configuration of these seeds might differ for different jet numbers. Therefor, the additional factor influencing the ratios of exclusive jet cross sections is n dependent. To have a covered jet we need to have at least one uncovered jet[39]. Therefor the ratio between these two cannot go beyond 50%. Of course, for more densely packed configurations we get just more even smaller jets. Assuming that low multiplicities do not produce many overlap jets, but high multiplicities do, we use

$$f_n = \frac{a}{2}\left(1 + \frac{1}{\exp\left[\frac{n-b}{c}\right]+1}\right) \tag{1.42}$$

to describe how the ratio between full and partial jets changes. Here a, b, and c are constants. This distribution is inspired by a Fermi–Dirac statistic. The parameter b describes where the change from mostly full jets to mostly partial jets occurs. c tells us how abrupt this change happens. The parameter a helps us to absorb some of the uncertainties introduced due to the crude approximations, which guide our formula. The phase space suppression factor between jet ratios is then given by

$$\varphi_{\frac{n+1}{n}} = \frac{4-(n+1)R^2 f_{n+1}}{4-nR^2 f_n}. \tag{1.43}$$

More then one parton to start with

In reality we will have several partons as seed for the evolution equation. We simply need to multiply their respective generating functionals before taking the u derivative. This results just in a combinatorical problem. For Poisson configurations that is trivial. For staircase scaling we compute it in App. C.4. For i initial gluons, assuming staircase scaling *without* breaking terms, we find

$$R_{\frac{n+1}{n},i} = \frac{n+i}{n+1}(1-\Delta_g(e,\xi)). \tag{1.44}$$

For large n, which is exactly the region we are working in, we get pure staircase scaling again. Note that what large n is does depend on the particular process and its phase space boundaries. Therefor, staircase scaling is not dependent on the number of external legs of the matrix element we use to describe the hard interaction, but might sometimes only be visible in the high multiplicity region.

Monte Carlo study

We have computed both staircase as well as Poisson scaling from first principles in QCD using the generating functional formalism. This means that these features are resummed to all orders in perturbation theory at LL accuracy. Within their respective limits they are solid QCD statements. We also studied the most leading effects contributing to phase space suppression factors as well as the combinatorics for any given initial condition in the pure staircase case. We now carefully compare these analytic prescriptions with MC data. From Sec. 1.1 it should be clear that the generating function and the parton shower are two sides of one medal. We may call the generating function an analytic parton shower. A MC computation gives us the additional information of matrix element corrections to the lower jet bins and the full kinematic distribution. On the other hand it is a Markov chain and therefor a black box regarding the structure of its results. This means, that we expect the MC to reproduce our earlier results, but in addition we get a glimpse on the quality and validity of our assumptions. Additionally MC events open the path to translate our analytic findings to the hadron collider system later on.

We generate events of the form $e^+e^- \to q\bar{q} + ng$ with the Sherpa event generator [47] at $\sqrt{s} = 1000$ TeV. We chose this high center of mass energy to allow for good statistics in the high

[39] Assuming not to high multiplicities at the moment. We will do a proper MC check of all the computations done here in the next section.

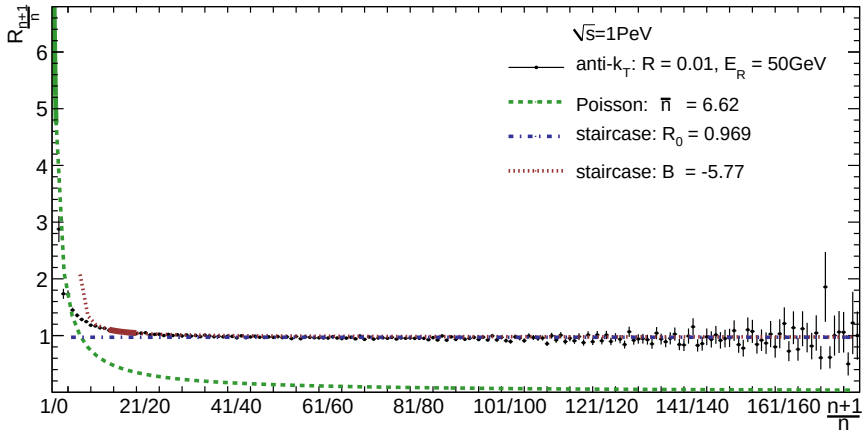

Figure 1.15: Exclusive jet cross section ratios in $e^+e^- \to q\bar{q} + ng$ with the e^+e^- anti$-k_T$ algorithm for $R = 0.01$ and $E_R = 50$GeV. The solid line corresponds to the bins used for the fit to eqs. (1.45) and (1.46). The dashed lines indicate the predicted values outside the region of fit.

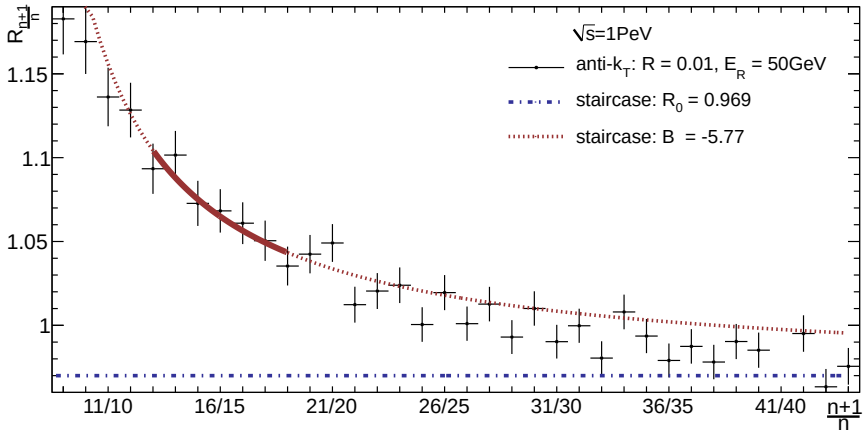

Figure 1.16: Exclusive jet cross section ratios in $e^+e^- \to q\bar{q}+ng$ with the e^+e^- k_T algorithm for different $R = 0.01$ and $E_R = 50$GeV. The solid line corresponds to the bins used for the fit to Eq. (1.45). The dashed lines indicate the predicted values outside the region of fit. Same as Fig. 1.15 but zoomed in the region used for fitting. Note the different scale on the x and y axis.

multiplicity regime and to avoid internal cut off scales. We compute the exclusive jet cross section ratios using the generalized e^+e^- anti$-k_T$ algorithm [38][40]. We chose this particular algorithm, because it produces circular shaped jets of half-opening angle R. This fact is useful if we like to study phase space effects in the light of Eq. (1.43). We then fit the formula

$$R_{\frac{n+1}{n}} = R_0 \times \left(1 + \frac{1}{B + (n+1)}\right) \tag{1.45}$$

to the tail of the jet cross section ratios. This will happen in the high multiplicity region and we do not have to take care of Eq. (1.44). We also show a fit to Poisson scaling for the first bins of the form

$$R_{\frac{n+1}{n}} = \frac{\bar{n}}{n+1}. \tag{1.46}$$

As we are studying a radiation phenomenon we adept the jet counting from [29] explained in the previous section and drop the first two jets as they belong to the hard event[41]. We expect staircase scaling only for the regime $n > \bar{n}$. To achieve both a large logarithm to check the Poisson hypothesis and a large multiplicity to check the staircase hypothesis we chose $R = 0.01$ and $E_R = 50$GeV.

[40] Note that at LL k_T, C/A, and anti$-k_T$ are equivalent [124].

[41] This means that experimentally our zeroth jet cross section has indeed two jets in the event.

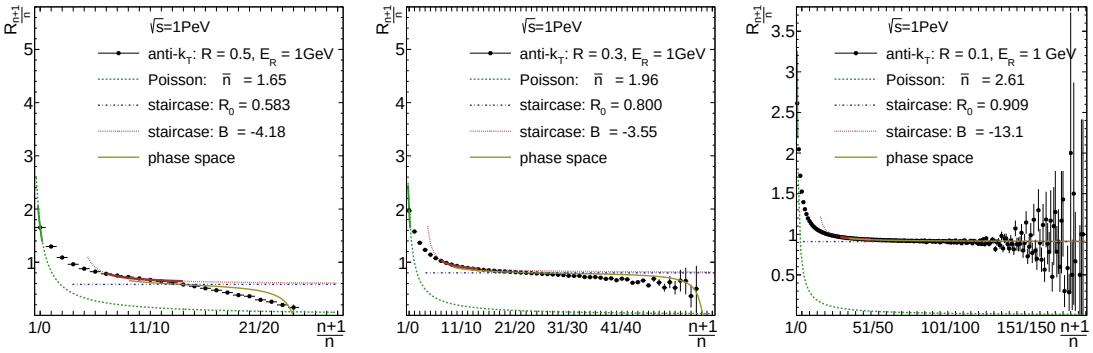

Figure 1.17: Exclusive jet cross section ratios in $e^+e^- \to q\bar{q} + ng$ for three different values of R. We fit Eq. (1.45) to the tail of the distribution (red curve) and multiply it with Eq. (1.43) (yellow curve).

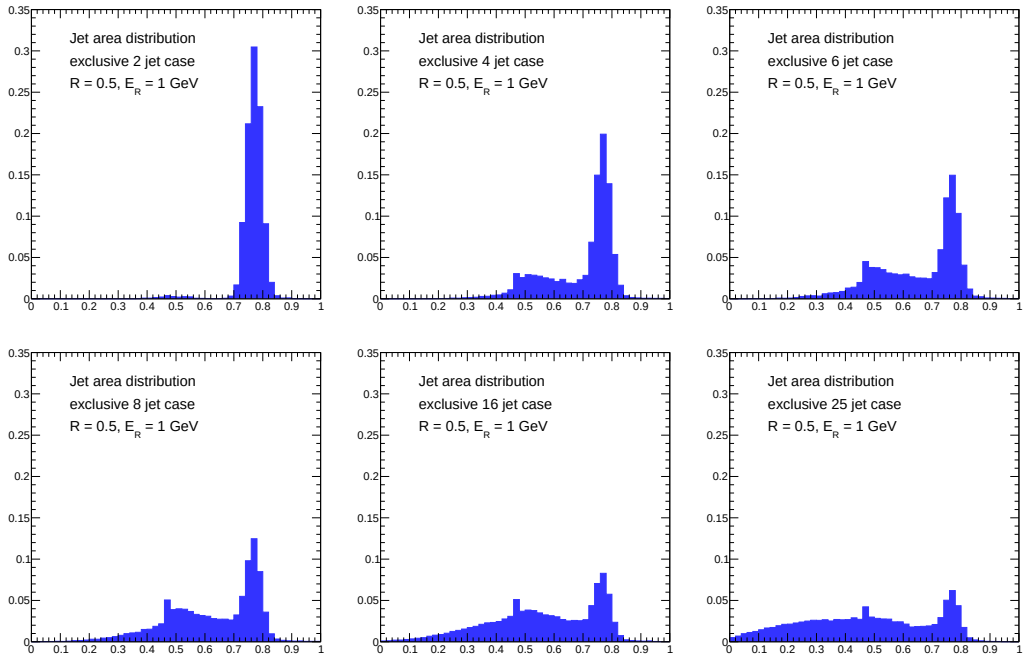

Figure 1.18: Jet area distribution at $R = 0.5$ ($A = 0.79$) and $E_R = 1$ GeV. Starting in the left upper corner we show the 2, 4, 6, 8, 16, and 25 jet exclusive case.

The small value for R allows for a large log and high jet multiplicities as well as the avoidance of phase space effects while the high value for E_R ensures that we do not get in trouble with MC internal shower cut-off. We show the cross section ratios in Fig. 1.15. We observe that both the Poisson and the staircase hypothesis fit the run of the curve very well within their expected regions of validity. The high value of R_0 is due to the small R we chose. The staircase pattern expands over the whole high multiplicity regime. A nice demonstration of the validity of Eq. (1.41) and the democratic scale nature of staircase scaling. Furthermore, we find that we only need to fit a few bins to Eq. (1.45) to correctly describe the distribution. It is worthwhile noting that the bins used to fit the staircase hypothesis do not fall into a region which would be qualified as staircase scaling, see Fig. 1.16. In fact we see the importance of the staircase breaking terms. They allow us to obtain information about the staircase tail of the distribution without the need to even see it. This knowledge can be useful in data driven background estimates. We can check the lower multiplicity bins for their SM compatibility. This is usually much easier as they accumulate more statistics. Then we use them to fit the QCD background to the staircase hypothesis. The value of R_0 and B is purely estimated by data this way and predicts the whole high multiplicity regime of the jet cross section ratios.

In the last paragraph we check for the existence of staircase scaling and the staircase breaking

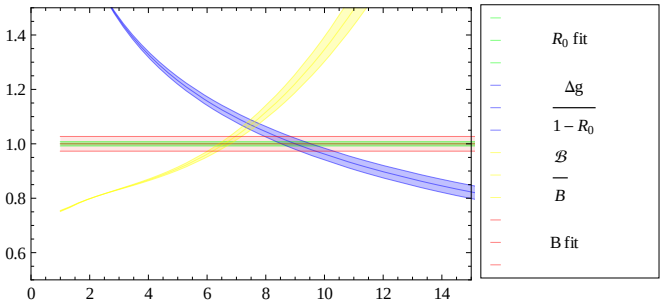

Figure 1.19: $\Delta_g(e)$ (blue) and $\mathcal{B}(e)$ (yellow) normalized to the fitted values R_0 (green) and B (red) found for $R = 0.5$ and $E_R = 50$GeV. The bands indicate either the parametric fit error or the variation due to different choices of $\alpha_s(M_Z)$.

terms. Therefor, we chose small R and moderate E_R. In usual collider studies values of $\mathcal{O}(0.10)$ are more realistic for R. With this value for the jet size we expect to see phase space suppression effects due to Eq. (1.43). This equation is purely geometric. To check its validity we now chose a very small $E_R = 1$GeV. This way we avoid phase space effects due to energy conservation. Using higher values for R we are still save from MC internal shower cut-off scales. We show the usual staircase fit for three different values of R $(0.5, 0.3, 0.1)$ in Fig. 1.17. We find that Eq. (1.43) describes the drop off of the distribution well. However, the shape is not described well. To understand why our simple picture of overlapping jets failed it is instructive to study the actual distribution of jet areas for a fixed multiplicity. In Fig 1.18 we show how the area distribution changes for increasing jet multiplicities. As expected we do find a peaky structure at 0.4 of the maximally possible area. However, we see that there are many jets with different areas. In the high multiplicity regime the area is nearly equally distributed over the whole available range. This explains why our phase space formula in Eq. (1.43) is not able to fully describe the phase space suppression. Understanding the effective area better could help to use jet scaling features in subjet studies, too.

Numerical quality of analytic formula

The analytic approach to staircase scaling has one unpleasant weakness. We use the dimensionless variable $e \equiv E/E_R$ in the limit where it approaches unity. In this limit we are able to find a closed solution to the generating function which produces staircase scaling. We are even able to write down the equation as a formal Taylor expansion. Using only the first expansion terms we compute a correction to the strict staircase scaling. Both, staircase scaling and the breaking term, we find realized in a MC study. The patterns appear universal, i.e their exact values depend on the exact choice of jet algorithm parameters, their existence does not. The problem with the expansion parameter e is that it is formally connected to two fixed values: the hard scale E and the jet definition E_R. By studying the limit we let e run. As we cannot change E_R by definition, this means we lose the connection to the hard scale of the event. So to speak, we expand from the soft limit away. On the one hand this is exactly what we want. Staircase scaling is a feature of democratic QCD evolution. There is only one scale: E_R. On the other hand we lose on the side of computability. We cannot tell what the correct E is. We do need a little bit of running, because otherwise there would be no jets at all. Therefor, we do not know, which e we need for the exact values of $\Delta_g(e, \xi)$ and $\mathcal{B}(e, \xi)$. However, we are still able to learn more about those quantities. The flat staircase tail defines the value $\Delta_g = 1 - R_0$. A priori the only free parameter is e as $\xi = 1$ for $q\bar{q}$ production [124]. Therefor, we can fit for e and plug it in $\mathcal{B}(e, \xi)$. First let us look at the characteristics of those two functions. In Fig. 1.19 we show the running of both, but normalized to fit the values found for $R = 0.5$ and $E_R = 50$ GeV. For a perfect prediction both curves would intersect at one. If we, for example, take $e = 7$ in this configuration we find Δ_g and $\mathcal{B}$ to be within 10% of the fitted values. For different parameter choices we find smaller and bigger uncertainties, but we get the order of magnitude right. However, the shown uncertainties are not the only ones entering our problem. As mentioned in App. C.3 we have no estimate for the uncertainty on $\mathcal{B}$ for $e > 1$. A situation we clearly face. Furthermore, the exact value of B, although indicated to be fitted within 10% strongly depends on the actual bins used for the fit. The fit of R_0 as well as the functional form of Δ_g are more reliable. There we are able to get a better statement about e.

Having an e of $\mathcal{O}(10)$ might seem to be far away from the $e \to 1$ limit. However, the full system is described by e of $\mathcal{O}(10^6)$. This shows, that we are clearly in the small e limit.

1.3.5 Hadron colliders

To translate the results from the last section to hadron colliders let us first study the consequences of the coordinate system used at hadron colliders. Hadron collider events obey only a cylindrical symmetry. Therefor, we use the pseudo rapidity η instead of the angle θ. This, however, changes which partons are assigned to which jet. For small distances $\Delta\theta = \theta_1 - \theta_2$ they are related in the following way

$$\begin{aligned} |\Delta\eta| &= \log \frac{\tan\frac{\theta_1}{2}}{\tan\frac{\theta_2}{2}} \\ &\approx \frac{\Delta\theta}{2/\tan(\theta/2)}. \end{aligned} \tag{1.47}$$

Here $\theta \approx \theta_1 \approx \theta_2$ is the direction of the jet. This means that we get less jets in the central region in the detector and more in the forward region (in the forward region the same jet will produce a bigger $\Delta\eta$ and will be divided in two hadron collider jets therefor). A further point which might affect the counted number of jets is the η acceptance of the calorimeter. This point is not invoked in our analysis so far. We study our MC sample with the combinations of R and E_R respectively $p_T^{\min}$ given in Tab. 1.5. For the hadron collider algorithm we add an η acceptance cut. However, the part of the sphere not covered by the detector is rather small. We do not expect big differences from this cut. In Fig. 1.20 to 1.22 we show the results for the e^+e^- anti$-k_T$ algorithm with no selection

R	0.1	0.3	0.5
$E_R/p_T^{\min}$[GeV]	10	50	90

Table 1.5: Values for R and lower jet scale used in the comparison between lepton and hadron collider jet algorithms.

cuts imposed in contrast to the corresponding hadron collider algorithm with an η acceptance cut. Each figure corresponds to a certain jet radius R. We find that Eq. (1.45) and (1.46) work out in the whole tested parameter space. For high E_R respectively $p_T^{\min}$ the high multiplicity regions show large scattering, but within statistical uncertainties the staircase hypothesis holds. The pattern we find in Fig. 1.15 and 1.16 are also valid. This strengthens the argument that staircase scaling as well as its breaking is not dependent on the specific jet algorithm or collider environment. In addition we observe, that indeed hadron collider coordinates produce more jets. This leads to an even more stable staircase scaling tail in all distributions. We find that the fitted values do change, however their relations and sizes do not.

The following two subsections "Generating functional for incoming hadrons" and "Parton density suppression" have already been published in [29] together with Erik Gerwick, Tilman Plehn, and Steffen Schumann. Note a change in conventions. The generating functional until now was expressed in terms of e and ξ. In the following sections it will (again) be expressed in terms of scales μ. This has no impact on the results as different evolution variables are equivalent in the collinear limit. However the notation of factorization scale might be a rather unfortunate choice compared to the convention of scales used so far. In the following section we call the hard scale, where the matrix element lives Q^2. The scale where the parton shower stops and additional initial state radiation is resumed by the DGLAP equation will be denoted by μ_F. This yields the following mapping:

$$\begin{aligned} \text{old convention} &\to \text{ new convention}, \\ \mu_0 &\to \mu_F, \\ \mu_F &\to Q^2. \end{aligned}$$

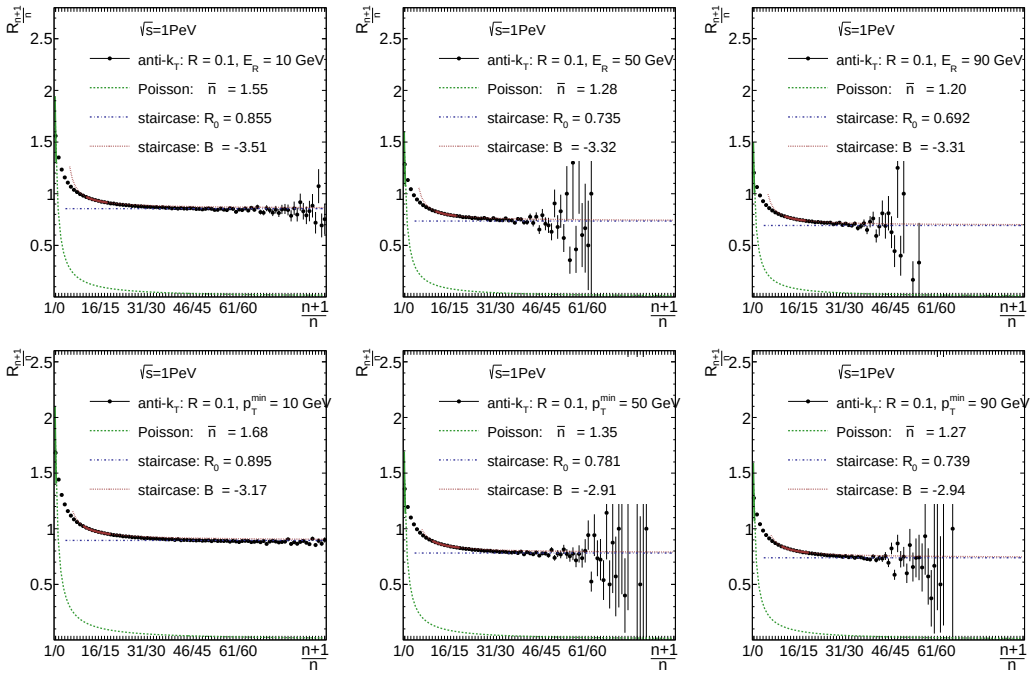

Figure 1.20: Exclusive jet cross section ratios in $e^+e^- \to q\bar{q} + ng$. Upper panels use the e^+e^- anti$-k_T$ algorithm. Lower panels use the hadron collider anti-k_T algorithm. We fix $R = 0.1$ and have from left to right E_R respectively $p_T^{\min}$: 10, 50, and 90 GeV. The solid lines correspond to the bins used for the fit to Eq. (1.45) and (1.46). The dashed lines indicate the predicted values outside the region of fit.

Generating functional for incoming hadrons

The basis of the QCD treatment of hadron collider physics is collinear factorization which allows us to employ the generating functional method [98, 99, 101]. Before we can apply any of this to jet counting we need to clarify our choice of the factorization scale μ_F in exclusive n_{jet} rates, i.e. in the presence of a jet-counting or jet-veto scale p_V. The resummation properties of the DGLAP equation identify the combined renormalization and factorization scale with a collinear cutoff below which initial state splittings are unresolved and influence only the functional dependence on the partonic energy fraction x. Because we are interested in radiated jets with $p_T \geq p_V$ we identify the factorization and the jet-veto scale, i.e. $\mu_F \equiv p_V$. Note that this choice furthermore avoids generating additional finite though potentially large logarithms in the ratio μ_F/p_V [103].

Symbolically, going from final state radiation in e^+e^- collisions to deep inelastic scattering (DIS) with initial state radiation and parton densities we replace the two generating functionals, distinguishing time-like from space-like splittings,

$$\Phi_q(Q^2, p_V^2) \times \Phi_{\bar{q}}(Q^2, p_V^2) \quad \to \quad \Phi_{q/\bar{q}}(Q^2, p_V^2) \times \mathcal{Z}_{q/\bar{q}}(x, Q^2, p_V^2). \tag{1.48}$$

In the original DIS context all scales are defined in terms of the e^+e^- Durham algorithm [101], most notably the hard scale Q and $p_V \equiv \mu_F$ as well as the softer resolution scale $Q_0 \leq \mu_F$. We identify all three relevant scales $Q_0 = \mu_F = p_V$. For the DIS analysis this corresponds to not further resolving the original macro-jets which define the separation of resolved jets and beam jets [101]. Again, this choice omits potentially large finite scale logarithms in our perturbative treatment.

We also introduce an explicit x dependence in the generating functional for incoming partons as it is clear that PDF effects alter the possibility to radiate jets. Each emission takes away an energy fraction $1 - z$ of the emitter; the x value has to change correspondingly and splitting between different partons needs to be taken into account. From the factorization theorem we know that PDFs and partonic cross-sections also factorize at the generating functional level,

$$\mathcal{Z}_a(x, Q^2, p_V^2) = \sum_b \int_x^1 \frac{dz}{z} \; f_b\left(\frac{x}{z}, p_V^2\right) \mathcal{Z}_a^b(z, Q^2, p_V^2) \; . \tag{1.49}$$

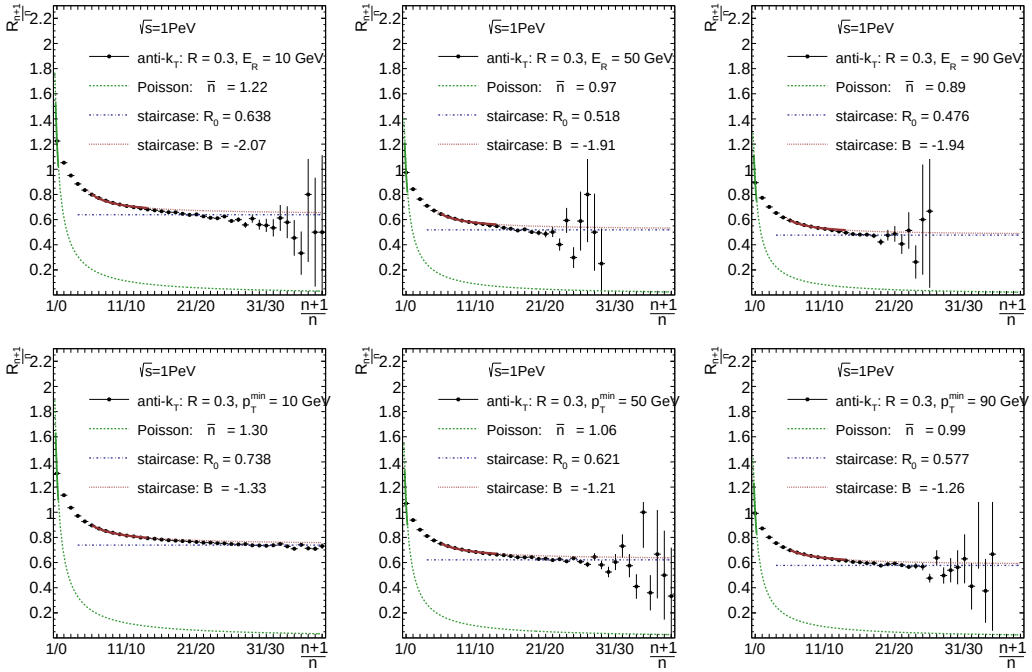

Figure 1.21: Exclusive jet cross section ratios in $e^+e^- \to q\bar{q} + ng$. Upper panels use the e^+e^- anti$-k_T$ algorithm. Lower panels use the hadron collider anti-k_T algorithm. We fix $R = 0.3$ and have from left to right E_R respectively $p_T^{\min}$: 10, 50, and 90 GeV. The solid lines correspond to the bins used for the fit to Eq. (1.45) and (1.46). The dashed lines indicate the predicted values outside the region of fit.

The parton densities we consistently evaluate at the scale p_V. This way, logarithmically enhanced parton splittings above p_V are described by the partonic generating functional $\mathcal{Z}_a^b$. For the generating functional in DIS we start with a time-like generating functional for a single (anti-)quark and weight it with the proper electromagnetic coupling [101]

$$\Phi_{\text{DIS}} = \sum_a e_a^2 \; \Phi_a(Q^2, p_V^2) \; \mathcal{Z}_a(x, Q^2, p_V^2). \tag{1.50}$$

The partonic cross-sections and jet evolution are the same for the quark and the anti-quark, but the PDFs are different. In DIS the final state kinematics fix x. Additional jets radiated off the incoming parton imply that in this case we probe higher x values as given by the convolution in Eq. 1.49.

The task is to find the evolution equations for $\mathcal{Z}_a^b$. To leading logarithm (LL) this turns out to be relatively simple. In the soft and collinear limit [98, 99] the eikonal approximation implies $z \approx 1$. Furthermore, the $g \to q\bar{q}$ splitting is logarithmically suppressed compared to the other splittings, so we can neglect it. Under these two LL assumptions the evolution equation and the corresponding generating function in 1.49 read [101]

$$\begin{aligned} \mathcal{Z}_a^b(z, Q^2, p_V^2) &= \delta(1-z) \; \delta_a^b \; \Phi_a(Q^2, p_V^2) \\ \mathcal{Z}_a(x, Q^2, p_V^2) &= f_a(x, p_V^2) \; \Phi_a(Q^2, p_V^2) \; . \end{aligned} \tag{1.51}$$

The PDF effects and the jet generating function factorize in x, so we can treat them independently. In general, Φ_a is a two-scale generating functional [101]. Because we identify $Q_0 = p_V$ the second scale is suppressed and its evolution equation is almost the same as in the e^+e^- case. To leading logarithm we find

$$\Phi_a(Q^2, p_V^2) = \exp\left[\int_{p_V^2}^{Q^2} dt \; \Gamma_a(Q^2, t) \; \Big(\Phi_g(t, p_V^2) - 1 \Big) \right] . \tag{1.52}$$

Compared to Eq. 1.12 the factor u in front of the exponential is missing. The reason is that we cannot resolve a jet if there is not at least one space like splitting. The hard parton cannot

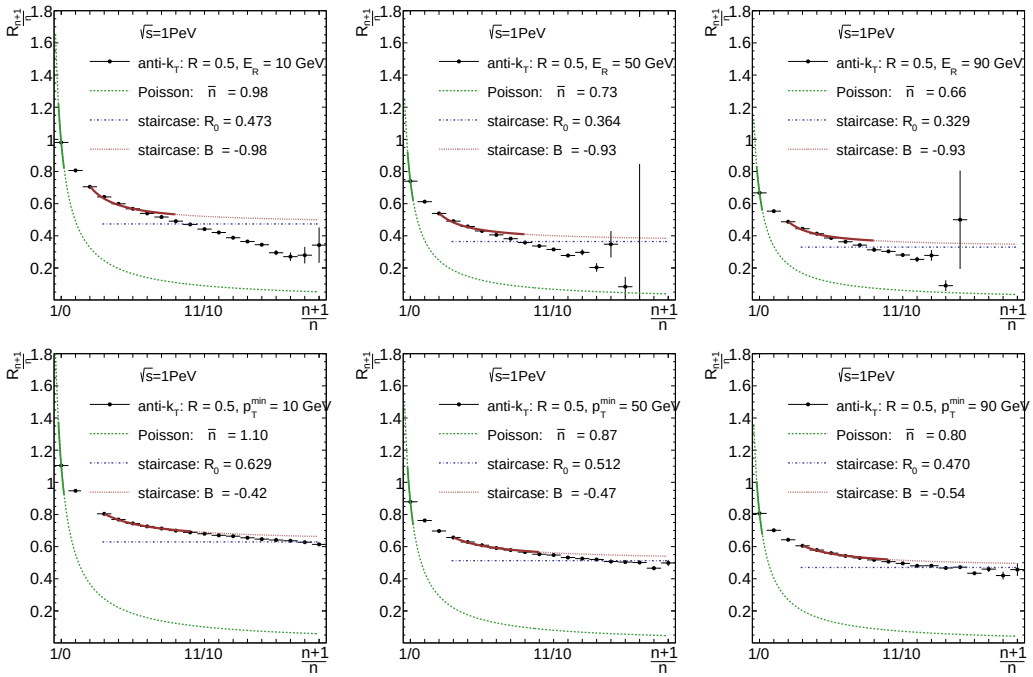

Figure 1.22: Exclusive jet cross section ratios in $e^+e^- \to q\bar{q} + ng$. Upper panels use the e^+e^- anti$-k_T$ algorithm. Lower panels use the hadron collider anti-k_T algorithm. We fix $R = 0.5$ and have from left to right E_R respectively $p_T^{\min}$: 10, 50, and 90 GeV. The solid lines correspond to the bins used for the fit to Eq. (1.45) and (1.46). The dashed lines indicate the predicted values outside the region of fit.

produce a final state jet, so we always find the normalization condition $\Phi_a(p_V^2, p_V^2) \equiv 1$. The further evolution of emitted partons we describe with the time-like functional of 1.12.

Moving on to Drell-Yan production with two incoming partons we need to replace the generating functionals, symbolically written, to

$$\mathcal{Z}_{q/\bar{q}}(x_a, Q^2, p_V^2) \times \mathcal{Z}_{\bar{q}/q}(x_b, Q^2, p_V^2) \; . \tag{1.53}$$

Thus, we replace the remaining time-like generating functional with a space-like generating functional to describe two incoming partons. A major complication is that the final state phase space does not fix $x_{a,b}$ anymore. Instead, we have to integrate over their allowed ranges and find the generating functional for the Drell-Yan process,

$$\begin{aligned}
\Phi_{\text{Drell-Yan}} &= \sum_{a,b} \int dx_a dx_b \; \mathcal{Z}_a(x_a, Q^2, p_V^2) \; \mathcal{Z}_b(x_b, Q^2, p_V^2) \\
&\approx \sum_{a,b} \int dx_a dx_b \; f_a(x_a, p_V^2) f_b(x_b, p_V^2) \; \Phi_a(Q^2, p_V^2) \Phi_b(Q^2, p_V^2) \\
&\approx \sum_{a,b} f_a(x^{(0)}, p_V^2) \Phi_a(Q^2, p_V^2) \; f_b(x^{(0)}, p_V^2) \Phi_b(Q^2, p_V^2) \; .
\end{aligned} \tag{1.54}$$

From this generating functional we can derive the individual n-jet rates. For the second line of Eq. 1.54 we use the leading logarithmic approximation as in the DIS case. To arrive at the third line we replace the variable $x_{a,b}$ values by a typical partonic energy scale $x^{(0)}$. For typical hadron colliders processes we assume this value to be close to threshold and equal for the two incoming partons. The argument u which generates the different n-jet rates is carried only by the generating functionals $\Phi_{a,b}(Q^2, p_V^2)$. Starting with two generating functionals for the two initial state particles, hard jet radiation with $p_T > p_V$ indeed factorizes from a PDF factor.

One apparent contradiction related to the PDF kinematics we need to resolve. On the one hand, in Eq. 1.51 the eikonal approximation allows us to set $z \approx 1$, which means that the entire energy dependence is encoded in the PDF factor. On the other hand, each resolved jet requires a

finite $p_T > p_V$. Hence, the integration range for $x_{a,b}$ is determined by the partonic n-jet process and $x^{(0)}$ implicitly depends on u. This implicit dependence we have to account for by hand. In particular for parton density regimes which increase towards small x the majority of multi-jet events at the LHC are produced at threshold. The threshold value for any of the n-jet production rates we denote as $x^{(n)}$, leading us to the modified factorized form

$$\Phi_{\text{Drell-Yan}} = \sum_{a,b} f_a(x^{(n)}, p_V^2)\Phi_a(Q^2, p_V^2)\; f_b(x^{(n)}, p_V^2)\Phi_b(Q^2, p_V^2)\;. \tag{1.55}$$

We emphasize that n is determined *a posteriori* upon differentiation with respect to u, so is presented for illustrative purposes only. 1.55 means that to leading logarithm the jet radiation pattern in the Drell-Yan case is the same as in $e^+e^- \to$ jets processes, modulo explicit PDF factors estimated using an n-dependent threshold kinematics. A similar approach can account for energy momentum conservation in soft-gluon resummation [94]. This way we leave the LL evolution of jets untouched and instead shift the x value in the PDFs to account for additional jets. All our findings from Sec. 1.3 we can immediately apply, once we understand the PDF correction factor in the next section.

Parton density suppression

In Sec. 1.3 we have learned that to leading logarithmic accuracy the effects of the parton densities and jet emission factorize. For large jet multiplicities this explains the observed staircase scaling at hadron colliders [95, 138]. Parton densities contribute to this effect in particular at low multiplicities. When increasing the jet multiplicity the typical partonic energy fractions x probed by the partonic process increase as well. The relative increase in x is largest for low jet multiplicities.

In terms of the assumed threshold kinematics adding a jet with finite transverse momentum implies $x^{(n+1)} > x^{(n)}$. To compute the relative cost of producing an additional jet we estimate the ratio of PDF values evaluated at $x^{(n)}$ and $x^{(n+1)}$ as a function of the number of extra jets n. In effect this is the discretized second derivative with respect to x. For hadron collider processes involving two parton densities $f(x,Q)$ we define the PDF correction factor to the ratio of successive jet ratios $R_{(n+1)/n}/R_{(n+2)/(n+1)}$ to be

$$B_n \;=\; \left| \frac{\dfrac{f(x^{(n+1)},Q)}{f(x^{(n)},Q)}}{\dfrac{f(x^{(n+2)},Q)}{f(x^{(n+1)},Q)}} \right|^2 \;. \tag{1.56}$$

The square in the definition of B_n reflects the two PDFs in hadron collisions. If for example the partonic ratio of two successive jet ratios is $R_{(n+1)/n}/R_{(n+2)/(n+1)} \sim c$ then the proper hadronic ratio becomes $B_n c$. We fix Q for simplicity, but this only mildly affects our results.

The main effects are, first, that $B_n < 1$ in most cases. This way PDF effects suppress the lower multiplicity ratios $R_{(n+1)/n}$. For large jet multiplicities the relative impact of yet another jet becomes small, $B_n \to 1$. The hadronic initial-state effect on the jet scaling disappears and we are back to the staircase pattern. Second, the PDF effect is largest for the steep gluon densities, as compared to the flatter valence quarks. Finally, allowing for variable Q the PDF values $f(x,Q)$ increase (decrease) with Q for low (high) x, with a cross-over point around $x \sim 0.1$. For small x values the initial state evolution then suppresses jet ratios at high multiplicity or large Q^2.

What we are most interested in are PDF effects for the Drell-Yan process at lower multiplicities. We consider the threshold values $x^{(n)}$, for example for producing an on-shell Z-boson and one additional jet,

$$x^{(1)} = \frac{\sqrt{m_Z^2 + 2\,(p_T\sqrt{p_T^2+m_Z^2}+p_T^2)}}{2\,E_{\text{beam}}}\;. \tag{1.57}$$

where $E_{\text{beam}} = 3500$ GeV for the LHC in 2011. Comparing $x^{(1)}$ with $x^{(0)} \approx m_Z/(2E_{\text{beam}})$ shows a sizeable shift. For the two-jet threshold $x^{(2)}$ two limiting cases are either the additional jet adding merely $p_T/(2E_{\text{beam}})$ to $x^{(1)}$ or two approximately collinear jets recoiling against a hard Z. The variation between these two cases estimates the uncertainty on our method which can be generalized straightforwardly to the n-jet final state.

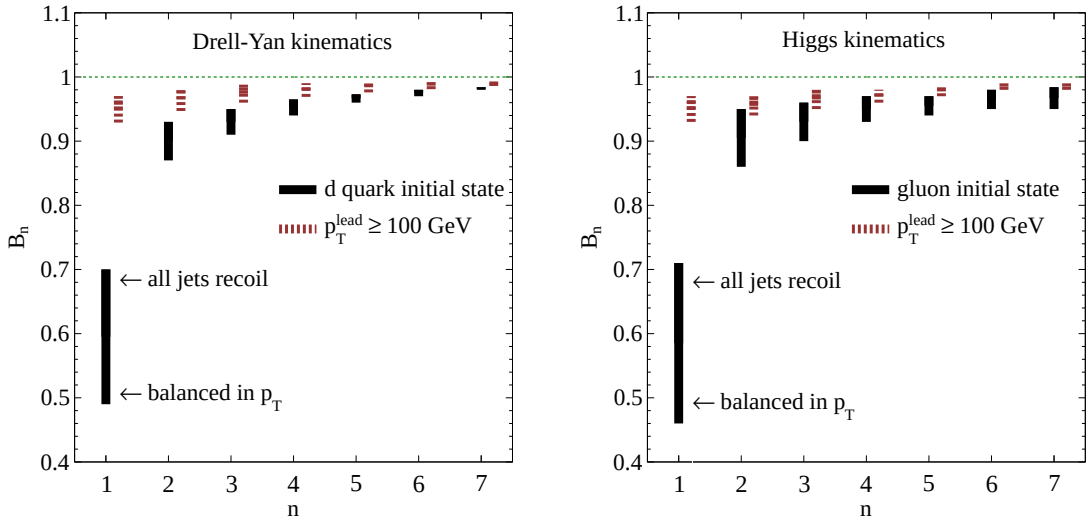

Figure 1.23: Left panel: estimated PDF suppression for inclusive (solid) and jet-associated (dashed, $p_T^{\text{lead}} \geq 100$ GeV) Drell-Yan kinematics. We assume an initial state with d-quarks only. Right panel: same for Higgs production in gluon fusion with $m_H = 125$ GeV. The uncertainty encompasses two representative kinematical limits of the multi-jet final state, described in the text.

In the left panel of Fig. 1.23 we display B_n for the estimated Drell-Yan kinematics, assuming each jet has transverse momentum $p_T = p_V = 30$ GeV. The effect on the first jet ratios is large, but quickly diminishes towards higher n. We also see that if we require a leading jet with large transverse momentum, $p_T^{\text{lead}} \geq 100$ GeV, we move to sufficiently high x such that additional jet ratios are unaffected by the PDF effect. It is reassuring to see that if we combine the PDF suppression of $R_{1/0}$ (0.46 - 0.65) with the C_A enhancement of $R_{2/1}$ (1.36) and assume an original Poisson scaling we find $R_{1/0}/R_{2/1} = (0.67 - 0.95)$, in nice agreement with ATLAS data [96]. This beautifully illustrates that staircase scaling at large multiplicities can be derived from first principles QCD while for small multiplicities it is something like a sweet spot.

As an additional check, we present the PDF suppression in gluon-fusion Higgs production in the right panel of Fig. 1.23. We assume $m_H = 125$ GeV, ignore flavor changes and consider jets with $p_T = p_V = 30$ GeV. The gluon PDF drops more rapidly for increasing x, inducing a large PDF suppression. On the other hand, the increasing energy of the core process as compared to the Drell-Yan process slightly decreases the effect. The combination of the two gives remarkably similar results to the Drell-Yan process.

Jets in Action

In the previous chapter we have shown that we can understand the jet number spectrum from first principle QCD. We already saw that multi jet observables are powerful tools to new physics, but suffer from theoretical uncertainties. In this chapter we put this knowledge to use. We first show how we can use scaling features as tool for data driven background estimation. Fixing the rates from experiment and predicting the shape of the jet spectrum from theory allows us to trust in multi jet observables. First we show in Sec. 2.1, that we can use the form of the jet spectrum as a calibration for interesting backgrounds. We show how staircase scaling as well as Poisson scaling can be used to understand this distribution. In Sec. 2.2 we show what we gain by including jets to all orders into a Higgs search instead of vetoing them.

Sec. 2.1 has already been published in [28] together with Christoph Englert, Steffen Schumann and Tilman Plehn. Also Sec. 2.2 has already been published in [33] together with Catherine Bernaciak, Bruce Mellado, Tilman Plehn, and Xifeng Ruan. However, we present an additional section at the end studying the source of the improvement in more detail.

2.1 Studying backgrounds from data: photon plus jets

2.1.1 Introduction

After testing and reproducing many interesting aspects of the Standard Model at the LHC, the focus of the ATLAS and CMS collaborations is rapidly moving toward searches for Higgs particles [145] or physics beyond the Standard Model [53]. The production rates for any of these search channels are small, for example compared to W/Z+jets or top pair production, channels which constitute their main backgrounds. Traditionally, at hadron colliders we have relied on the appearance of leptons, photons or missing transverse energy to point us to interesting new physics processes. In this approach QCD effects and jet production are either ignored or considered a nuisance.

Starting with the suggested searches for a light Higgs boson in weak boson fusion [125], this attitude has changed; this search shows how the QCD structure of signal events can be turned into a powerful handle to reject large backgrounds. The key analysis tool are (central) jet vetos [126, 127], which for example suppress QCD-initiated W/Z+jets events or hadronic top pair production. Implicitly, this approach is adopted in Higgs searches for example in the $H \to WW$ or $H \to \gamma\gamma$ channels, when those searches are divided into 0-jet, 1-jet and soon 2-jet strategies [145].

In a similar spirit, searches for example for supersymmetry benefit from the measurement of the number of jets which includes information on the color structure of the new heavy states [27]; the only caveat is that we need to carefully separate decay jets from QCD jet radiation associated with hard processes [27, 129]. What is missing for all such analyses is an experimentally established and theoretically sound link between choosing n-jet samples for an analysis and a systematic study of the corresponding n_{jets} distribution for signal and background processes [127]. A dedicated study along this line would map out the behavior of exclusive n_{jets} distributions after different cuts, understand its basic features, and quantify the notorious theory uncertainties associated with jet counting. As it will turn out, the production of a photon in association with QCD jets is a perfect basis for such studies.

At hadron colliders we know two fundamentally different scaling patterns, both of which have been observed experimentally. From text book quantum field theory we know that successive photon radiation off a hard electron — as well as successive gluon radiation off a hard quark — follows a Poisson pattern for the *exclusive number* of photons or gluons [22]. This pattern corresponds to a simple probabilistic picture of successive independent splitting. The splitting probability is linked to the coupling constant, the color factor, the form of the splitting kernel, and a scale logarithm. The problem with this Poisson scaling pattern is that since UA1/UA2 it has never

been observed for inclusive production processes at hadron colliders. Instead, for many processes we find staircase scaling, namely a constant ratio of exclusive n-jet rates $\sigma_{n+1}/\sigma_n = R$ [131, 132].

The description of exclusive jet rates is fundamentally at odds with our description of QCD at hadron colliders. Parton densities obeying the DGLAP equation [135] resum collinear logarithms and absorb the corresponding infrared divergences. As a consequence, any computation based on such parton densities is jet inclusive, i.e. it allows for an unspecified number of collinear jets radiated off the incoming partons. Strictly following the DGLAP approximation, these perfectly collinear jets are not observable at the LHC. However, the assumption of perfect collinearity is modified by the initial state parton shower, which redistributes the dominantly collinear jet radiation into the physical phase space. Evaluating exclusive event samples with exactly n jets is theoretically limited by the precision of the parton shower description, including its obvious breakdown for sizeable transverse momenta [136]. This is the reason why in the past exclusive n_{jets} distributions could rarely be exploited to compare collider data to QCD predictions. However, as described above, current LHC analyses force us to overcome this limitation and study exclusive n_{jets} distributions (including cuts corresponding to jet veto survival probabilities) starting from perturbative QCD.

Matching of a hard QCD matrix element with a collinear parton shower [45, 46, 60, 100] allows us to reliably simulate and study jet scaling patterns from perturbative QCD. It does not require a fundamental re-organization of QCD perturbation theory [137] but simply relies on the proper phase space simulation of collinear logarithms. This way it does not only include the radiation of one or two hard jets, as correctly described by fixed-order QCD calculations, but any number of radiated jets including high multiplicities obviously well described by the parton shower. We use Sherpa [47] with its Ckkw [45] matching scheme to describe radiation of up to seven jets with high precision. Typically, we check an additional two more jets for unexpected features, but with correspondingly reduced statistics.

As we will show in this paper, a particularly promising channel to measure exclusive jet rates and compare them to QCD predictions is hard photon production in association with jets. The cross section is large enough to already have enough data to not only validate Monte Carlos, but also systematically study staircase and other scaling patterns. We will show how to define and extract different kinematical regimes of the photon to compare various hypotheses of QCD radiation with data. Poisson and staircase scaling are the two basic structures which allow us to understand the performance of a central jet veto as a background rejection tool and which properly define the crucial n_{jets} and effective mass distribution in new physics and Higgs searches [126, 127, 145].

2.1.2 Staircase scaling

Before we start using QCD structures of events with heavy states, we study the n_{jets} distribution in Standard Model processes. This includes pure QCD jets production [27, 138], W/Z production with jets [131, 132, 138], and possibly our new γ+jets channel. For total cross sections such processes have for many years shown a staircase scaling pattern: if we evaluate the ratios of historically jet-inclusive n-jet cross sections, we find

$$\hat{R}_{(n+1)/n} = \frac{\hat{\sigma}_{n+1}}{\hat{\sigma}_n} \equiv \hat{R}\,. \qquad \text{(jet-inclusive)} \tag{2.58}$$

It turns out that we can equivalently formulate this condition in terms of inclusive ($\hat{\sigma}$) and exclusive (σ) numbers of jets. Correspondingly defining

$$R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} \equiv R \qquad \text{(jet-exclusive)} \tag{2.59}$$

for the exclusive rates, the resulting ratios are identical [27],

$$\hat{R} = \frac{\sum_{j=n+1} \sigma_j}{\sigma_n + \sum_{j=n+1} \sigma_j} = \frac{\sigma_{n+1} \sum_{j=0}^{\infty} R^j}{\sigma_n + \sigma_{n+1} \sum_{j=0}^{\infty} R^j} = \frac{R\sigma_n \dfrac{1}{1-R}}{\sigma_n + R\sigma_n \dfrac{1}{1-R}} = R\,. \tag{2.60}$$

This way the merits of perturbative QCD and its perturbative predictions directly translate to the jet-exclusive final states in a well-defined approach where lower multiplicities are utilized to constrain the higher ones.

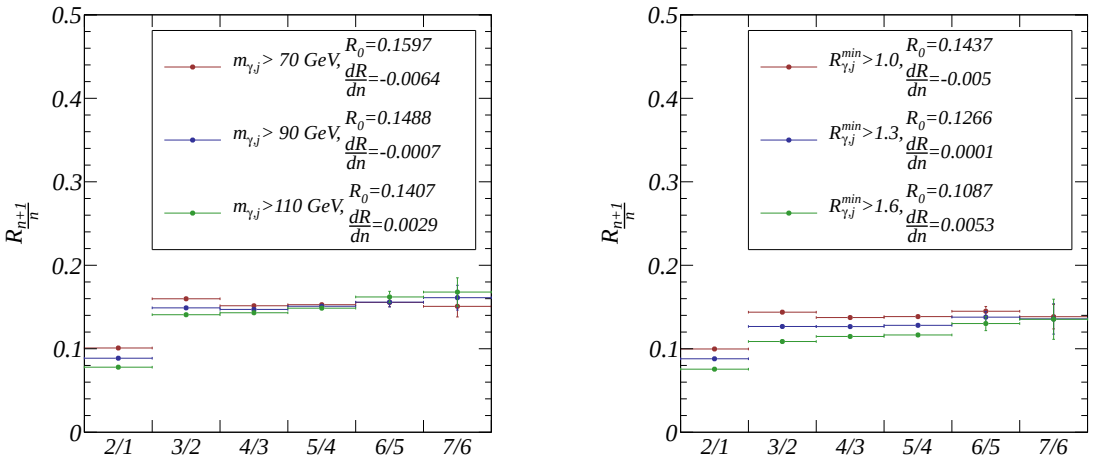

Figure 2.24: Two scenarios establishing staircase scaling for γ+jets production at the 7 TeV LHC. Left: invariant mass criterion of $m_{\gamma j} > 70,\ 90, 110$ GeV for each jet. Right: geometric separation between the photon and each of the jets. The extracted values for R_0 and dR/dn are defined in Eq.(2.62). The error bars correspond to our numerics with $1.6 \cdot 10^7$ events.

While a proper analytical derivation of this feature from first principles is still missing, there is no doubt that this feature exists in data [138, 139] and can be reproduced using matrix element and parton shower merging [27]. Only using jet radiation via parton shower this scaling feature is not matched as well, which is expected given that such inclusive processes do not offer a hard scale in relation to which we can define collinear radiation. For W+jets production it has been shown that fixed order QCD corrections stabilize the observed staircase pattern [70].

Unlike the other cases mentioned above, at first sight photon production does not posses an obvious jet scaling behavior. It only occurs once we include strong separation cuts between the photon and each of the jets, effectively removing any logarithmic enhancement linked to QED photon radiation. Moreover, as we will show in this paper, different basic cuts can easily induce different scaling patterns.

For our simulation we rely on Sherpa v1.3.0 [47] and its Ckkw matching up to five matrix element jets. We reconstruct jets using the anti-k_T algorithm from FastJet [38, 68] with $R_{\text{anti}-k_T} = 0.4$, which gives us a very moderate geometric separation of two jets. When dealing with photons in a QCD environment some familiar subtleties have to be considered [140, 141]: a photon can arise from non-perturbative fragmentation. Those photons are not useful in our case where we attempt to define a hard and identifiable core process through the photon. Therefore, we opt for a solid photon isolation. A naive hard cut e.g. on the jet-photon R distance limits the phase space of soft gluon emission and is infrared unsafe. We instead define an isolated photon through a hadronic energy deposit of less than 10% of $p_{T,\gamma}$ in a cone of size $R < 0.4$ [140]. If this criterion is not met the photon candidate is pushed into the jet finding algorithm.

For reconstructed jets and photons in this section we then require

$$p_{T,\gamma} > 50 \ \text{GeV}, \quad p_{T,j} > 50 \ \text{GeV}, \qquad |\eta_\gamma| < 2.5, \quad |y_j| < 4.5\,, \tag{2.61}$$

where η and y denotes the pseudo-rapidity and rapidity. These cuts are very inclusive and democratic, so we can expect to observe the staircase scaling behavior known for pure QCD jets.

It is known experimentally that photon plus jets events fulfilling Eq.(2.61) alone do not show any kind of simple jet scaling behavior. What we are still missing is the crucial photon-jet separation criterion. In Fig. 2.24 we show two sets of n_{jets} distributions for different separation criteria. Inspired by the W/Z+jets analysis we can define a wide photon-jet separation in terms of the invariant mass. In Fig. 2.24 we find that almost prefect staircase scaling appears for minimal values of $m_{\gamma j} \gtrsim m_Z$, with a very slight degradation for alternative mass scales. The corresponding description in terms of a geometric separation leads to very similar results, but only once we require $R_{\gamma,j} > 1$. This value we can understand from the typical $m_{\gamma j}$ values combined with $p_T > 50$ GeV.

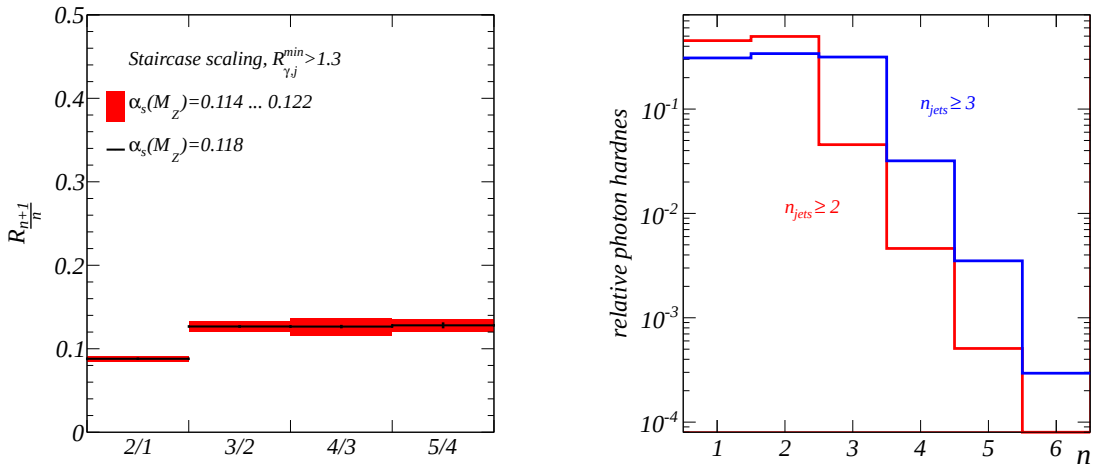

Figure 2.25: Left: effect of a consistent variation of $\alpha_s(M_z)$ on staircase scaling. Right: position of the photon in comparison to all jets, ordered by p_T. The two curves are for at least two-jet and at least three-jet events.

In both cases the first ratio $R_{2/1}$ is notorious [27], reflecting the fact that QCD scaling patterns require the definition of a stable hard process. Furthermore, the universal jet separation criterion imposed is not fully probed in the denominator, the γ+1 jet channel.

Theoretically, staircase scaling can be linked to the presence of the large gluon self coupling, i.e. the non-abelian structure of QCD. In the absence of any hard scale from our process, relatively hard jets are still mostly generated through initial state radiation (ISR). Our simulation confirms that the final state radiation (FSR) pattern of ISR jets and its combinatorial factors generates the large and democratic jet multiplicity which defines staircase scaling. In the pure Yang-Mills limit the number of Feynman diagrams scales exactly like the combinatorial factor from the bosonic final state phase space.

For hard photon production we start from $q\bar{q}$ annihilation. Adding a hard gluon radiated off one of the incoming quarks contributes to γ+1 jet production, but it is not the leading channel. As for many processes at the LHC, it is advantageous to start from a qg initial state and add a quark to the final state. The second jet defining $R_{2/1}$ will again be gluon radiation off the initial state. However, staircase scaling we know arises from splitting ISR, which for γ+2 jets would only be possible off the sub-leading $q\bar{q} \to \gamma$ process. This implies that the combinatorically enhanced non-abelian splitting starts from a suppressed parton luminosity, giving a relatively small value of $R_{2/1}$. In other words, the first entry $R_{2/1}$ links two different dominant initial states and hence cannot be described by a well defined hard process. Starting from two jets (or $R_{3/2}$) we can attribute the observed staircase scaling to splitting ISR gluons off the hard $qg \to \gamma q$ production process.

Setting aside the first entry we can test the quality of staircase scaling by fitting the form

$$R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} = R_0 + \frac{dR}{dn}\, n \,, \tag{2.62}$$

and determine the slope to compare it to our prediction $dR/dn = 0$. For all curves shown in Fig. 2.24 we find dR/dn in the $0.01 - 0.001$ range, essentially compatible with zero. The constant values R_0 range around 0.14, but with a small spread. In Sec. 2.1.4 we will contrast these values with the W/Z+jets cases [27, 131, 132]. While the definition of the hard process does play a role in determining R_0, in Sec. 2.1.3 we will see that the far dominant factor is $p_{T,j}^{\min}$ fixed in Eq.(2.61).

In the left panel of Fig. 2.25 we show the effect of α_s on the observed staircase scaling pattern. Between the largest and lowest values of α_s there is a 7% difference. The effect of this shift on R is correspondingly small. The reason is that α_s and for example the gluon parton densities are not independently extracted [71]. An increase in the value for the strong coupling is compensated

by a decrease in the corresponding gluon density, postponing the expected blow up of the theory uncertainty to larger n values than we can show in Fig. 2.25.

In the right panel of Fig. 2.25 we can check the reason why the first ratio $R_{2/1}$ is comparably small. The production of one hard jet with one hard photon is dominated by the partonic subprocess $qg \to q\gamma$. If there is no hard scale present in the partonic process the evolution of the incoming quark or gluon is dominated by few splittings. The large number of jets constituting staircase scaling arises through non-abelian splitting of ISR. Geometrically, for 2-jet events the quark-photon system typically recoils against the harder of the QCD jets. Only from the 3-jet configuration on we can simply split the ISR gluon, with a jet separation given by $R > 0.4$ according to the jet algorithm. Correspondingly, in the right panel of Fig. 2.25 we see that the photon is typically as hard as the hardest jets: if we require at least two jets the photon is the hardest or second hardest object in roughly half of the events each and the third hardest object in $\mathcal{O}(5\%)$ of all events. For at least three jets in the final state the photon is equally likely to be the first, second and third hardest object. In other words, while one of the jets might usually recoil against a relatively hard photon, the additional jets responsible for the staircase scaling pattern are relatively soft. This is a result of splitting hard ISR gluons.

2.1.3 Poisson scaling

According to field theory text book knowledge radiating massless gauge bosons off a hard fermion does not follow a staircase pattern [22]. Instead, the exclusive number of jets follows a Poisson distribution

$$\sigma_n = \sigma_0 \, \frac{e^{-\bar{n}} \, \bar{n}^n}{n!} \, , \tag{2.63}$$

with an expected $\bar{n}$ jets observed. For the exclusive scaling ratios this translates into

$$R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} = \frac{\bar{n}}{n+1} \, . \tag{2.64}$$

The assumptions entering the derivation of Poisson scaling are twofold: first, there should be one splitting function, for example the radiation of a photon or a gluon off a fermion. In the soft limit successive gauge boson radiation is automatically ordered by the emission angle. This way, we avoid combinatorial factors of the kind $n!$ from differently ordered emission in the numerator. The crucial factor $1/n!$ in Eq.(2.63) appears through the over-counting of the bosonic phase space. Poisson scaling is what one expects from a statistical point of view when we assign probabilities to statistically independent splittings. An example for this statistical treatment are Sudakov factors or collinear splitting probabilities following the DGLAP equation. The reason why solutions of the DGLAP equation show a Poisson behavior is that in its derivation we only take into account successive splittings of incoming partons on their way from the proton to the hard process. This is exactly what corresponds to a resummation of collinear logarithms and the removal of infrared divergences through the definition of scale dependent parton densities. For this reason, pure parton shower simulations tend to reproduce Poisson scaling much better than staircase scaling in the respective kinematic regime.

To force multi-jet scaling for example in association with a photon into such a Poisson regime we can follow this argument of the parton shower and the Sudakov factors. What we need is a well defined hard subprocess, e.g. the leading 2-particle γ-j_1 system which requires many successive splittings of the incoming partons.

Following the distance measures shown in Fig. 2.24 an obvious choice could be an increased value of $m_{\gamma j_1} \gg 100$ GeV. This is similar to weak boson fusion Higgs production, where the large invariant mass of the two tagging jets $m_{jj} > 600$ GeV induces a Poisson scaling of the Z+jets backgrounds [127]. However, we find that the cleanest Poisson distribution is induced by requiring a single hard jet, i.e. requiring

$$p_{T,\gamma} > 20 \;\; \text{GeV}, \quad p_{T,j} > 100, 20, 20, ... \;\; \text{GeV}, \qquad |\eta_\gamma| < 2.5, \quad |y_j| < 4.5 \, , \tag{2.65}$$

instead of Eq.(2.61). Generating the hard scale through the hardest jet is more efficient than asking for a hard photon, because according to our earlier argument a hard photon with $p_{T,\gamma} = 100 - 200$ GeV could easily recoil against several jets from splitting ISR. Requiring a hard average p_T for the jets would work as well, though.

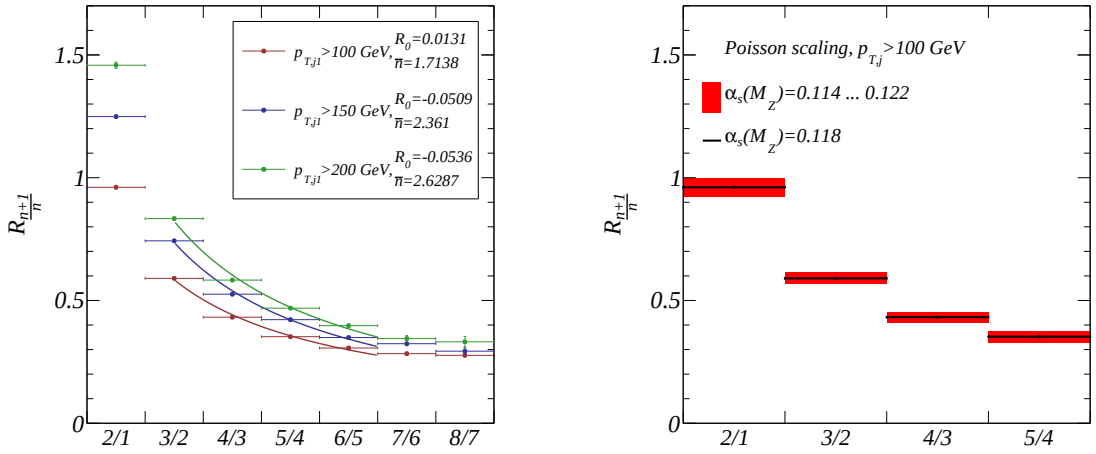

Figure 2.26: Poisson scaling for γ+jets production at the 7 TeV LHC. Left: different transverse momentum criteria for the leading jet, all other jets have $p_{T,j} > 20$ GeV. The extracted values for R_0 and $\bar{n}$ are defined in Eq.(2.66). The error bars correspond to our numerics with $2.1 \cdot 10^6$ events. Right: effect of a consistent variation of $\alpha_s(m_Z)$.

In Fig. 2.26 we see how enforcing a hard core process this way immediately changes the staircase scaling pattern into a Poisson distribution. Already for $p_{T,j_1} > 100$ GeV we see a clear deviation from any kind of staircase behavior provided we allow all other jets to be as soft as $p_{T,j} > 20$ GeV. For $p_{T,j_1} > 150$ GeV the leading $R_{2/1}$ ratio increases to values above unity, which means that in the exclusive n_{jets} distribution the maximum will move away from zero.

To test the quality of the Poisson description we fit the $R_{(n+1)/n}$ distribution which is expected to follow Eq.(2.64). If we allow for a deviation from the one-parameter Poisson shape of the kind

$$R_{(n+1)/n} = \frac{\bar{n}}{n+1} + R_0 \,, \tag{2.66}$$

R_0 is reminiscent of the staircase pattern Eq.(2.60) and should come out essentially zero while $\bar{n}$ is the only free parameter in the Poisson shape and gives the expected number of jets. As expected, the value of $\bar{n}$ increases for harder leading jets. Just as for the staircase scaling we do not include the first entry $R_{2/1}$ in the fit.

While it follows the basic expectation, namely becoming large and exceeding unity, $R_{2/1}$ does not fit the Poisson shape well. Again, this is due to the definition of the hard process with consistent incoming partons. Initial state radiation off the $\gamma + 0$-jet process involves two incoming quarks, with the typical color factor $C_F = 4/3$. Once the numerically dominant subprocess $qg \to \gamma q$ is established, on one leg we can expect ISR with a color factor $C_A = 3$. Hence, the 1-jet fraction is slightly too small and $R_{2/1}$ comes out too large. While this effect is much smaller than the mismatch of $R_{2/1}$ for staircase scaling, it is important to not include this bin into the fit to Eq.(2.66) because it would lead to obviously wrong best-fit values for R_0.

If we also exclude the high-multiplicity bins we find small values of $|R_0|$, closer to zero than to $R_{7/6} \sim R_{8/7}$. This is expected. Our argument for Poisson scaling rests on the impact of a large scaling logarithm which has to be generated by successive and ordered ISR. Beyond some point these successive splittings will stop feeling a large logarithm and we can expect to fall back onto a non-negligible staircase scaling.

For example in the case of Higgs production we know that the large-n limits of $R_{(n+1)/n}$ in the staircase and Poisson setups show hardly any difference. In Fig. 2.26 we estimate the staircase tail to be in the $R_0 \sim 0.3$ range. This is significantly different from the $R_0 \sim 0.15$ which we find in Fig. 2.24. The reason is simply the reduced p_T threshold of 20 GeV for the Poisson studies. In the right panel of Fig. 2.26 we again show the consistent variation of α_s. Clearly, the dependence is very small and does not affect the Poisson scaling feature.

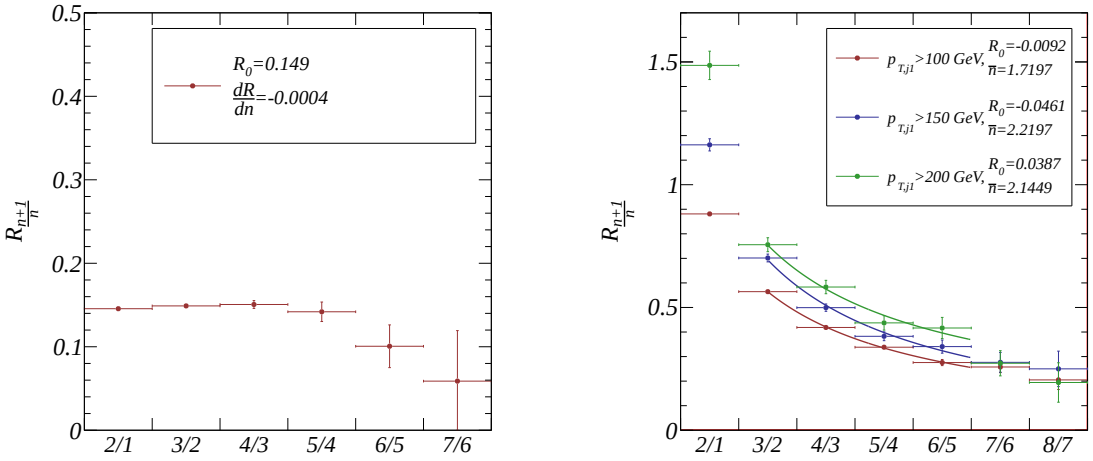

Figure 2.27: Scaling patterns for Z+jets production at the 7 TeV LHC. Left: staircase scaling for $p_{T,j} > 50$ GeV. Right: Poisson scaling for different transverse momenta of the hardest jet, all other jets have $p_{T,j} > 20$ GeV. The error bars correspond to our numerics with 10^7 events before cuts.

2.1.4 Massive gauge bosons

Originally, jet scaling studies have been established for W+jets events. Not surprisingly, Z+jets events behave qualitatively and quantitatively exactly the same [27]. The purpose of the first two Sections of this paper is to show that after strict jet-photon isolation we can see the same scaling patterns in photon production.

Provided that for well separated photons the non-existence of large logarithms between inherent mass scales leads to staircase scaling we expect γ+jets and Z+jets production to be very similar. In Fig. 2.27 we first see that based on the same jet cuts as in Eq.(2.61) Z production follows the same staircase pattern. The extracted value $R_0 = 0.149$ for $p_{T,j} > 50$ GeV is consistent with the literature [27] as well with our findings in Sec. 2.1.2.

For the Poisson regime the situation is slightly different. For the photons a cut on the leading jet of $p_{T,j_1} > 100$ GeV compared to a reduced general jet threshold of 20 GeV already induces a large enough logarithm. The Z mass in the final state could be expected to further enhance this scaling logarithm, even though it does not really translate into a collinear logarithm expressed in terms of $p_{T,j}$. In the right panel of Fig. 2.27 we find that the $\bar{n}$ values we extract from the Z case are very close to those for the photons, in particular taking into account the statistical uncertainties which affect the result for $p_{T,j_1} > 200$ GeV. The high-multiplicity staircase limit of the Poisson distribution for $R_0 \sim 0.20 - 0.25$ again is similar to the photon case.

From both figures there is little doubt that any result from γ+jets analyses at the LHC can be directly translated into Z+jets and hence W+jets [27] production. This holds not only on the qualitative but also on the quantitative level. The only difference between the two channels is that for photons the scaling patterns only appear once the photon is very well separated, controlling any additional QED logarithms which for the massive Z case do not play any role at these energies.

2.1.5 Outlook

Counting numbers of exclusive jets has many applications in LHC searches, implemented for example as distinct Higgs analyses for different n_{jets} values [145] or central jet vetos [125]. To apply such cuts we need to properly understand exclusive n_{jets} distributions experimentally and theoretically. For QCD this distribution is a challenge because historically it involves parton shower simulations. Multi-jet merging, for example using the CKKW scheme [45], allows us to for the first time study n_{jets} distributions including a free choice of kinematic cuts.

Unlike W/Z+jets and pure QCD jets production, the associated production of jets with a hard photon naively does not show simple scaling patterns. We show that once we require a widely separated photon we recover the usual staircase scaling $\sigma_{n+1}/\sigma_n = R_0$ for total cross sections [27, 131, 132].

Once we induce a large logarithm through kinematic cuts we see how the scaling pattern turns into a Poisson distribution for the exclusive number of jets. This is known for weak-boson-fusion cuts in Higgs production [127]. For our photon channel a transverse momentum cut on the leading jet (and not on the photon) works best. Only in the high-multiplicity regime the Poisson distribution turns into a staircase pattern.

Given our observations and the large available photon sample at LHC, this channel should be the perfect laboratory to study jet scaling, including a proper experimental and theoretical error analysis. A translation of photon measurements into W/Z+jets production should pose no problem and is of extreme importance when e.g. relating γ+jets to $(Z \to \nu\bar{\nu})$+jets as a background to new physics searches in the jets+missing energy channel [144]. This is just one example of how we should be able to significantly improve the status of many Higgs and new physics searches at the LHC once such general analyses are available.

We are grateful to Alex Tapper for pointing us to this channel. The simulations underlying this study have been performed in parts on bwGRiD, member of the German D-Grid initiative, funded by the Bundesministerium für Bildung und Forschung and the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg.

2.2 Learn more about the Higgs: Fox–Wolfram–Moments

2.2.1 Introduction

After the recent Higgs discovery by ATLAS and CMS [54,55,145], the careful and systematic study of Higgs properties is becoming a key research program at the LHC and a future linear collider [146]. The theoretical implications of the first fundamental scalar particle include many open questions, including the actual generation of a vacuum expectation value, the stability of its physical mass, or the link between the Higgs potential at the weak scale to high–scale structures [147]. In the language of quantum field theory we need to construct the weak–scale Higgs Lagrangian including the operator basis and the corresponding couplings [148].

At the LHC the weak boson fusion production channel (WBF) [53,149–152] plays an important role in answering some of these question, in particular once the LHC runs closer to its design energy. It allows us to directly probe the unitarization of $WW \to WW$ scattering and carries information on tree–level Higgs couplings with negligible impact of perturbative extensions of the Standard Model. Experimentally, two forward tagging jets are highly effective in reducing QCD backgrounds [153], which means that Higgs analyses in weak boson fusion typically benefit from a signal–to–background ratio around unity.

As an analysis tool utilizing the unique QCD structure of weak boson fusion we rely on a central jet veto [103, 136, 154–157]. It is based on the fact that we can generate large logarithms and increase central jet radiation in QCD backgrounds while leaving the jet activity in the signal at low level. This shift from staircase scaling of jets (with constant ratios between successive exclusive jet bins) in signal and background to staircase scaling in the signal and Poisson scaling in the background can be derived from first–principles QCD [154]. The resulting jet veto survival probabilities for the the QCD backgrounds can be measured in data. Their calculation from QCD is plagued with significant theory uncertainties which in turn will soon dominate the extraction of the Higgs couplings at the LHC [148]. In addition, a jet veto always removes a wealth of kinematic information carried by these jets, so the question arises whether the information from the jets recoiling against the Higgs cannot be used more efficiently.

To answer the question of how much information is encoded in the jet activity of Higgs candidate events we need to systematically study multi-jet kinematics. For example in flavor physics Fox–Wolfram moments (FWM) are an established tool to analyze such geometric patterns [158], but they have hardly been employed by the ATLAS and CMS collaborations. By construction, they are particularly well suited to study the geometry of tagging jets in weak boson fusion [159]. Dependent on the specific construction of their weights the moments can also be sensitive measures of the additional jet activity in an event. Ideally, they will enhance a central jet veto defined on a fixed phase space region to some kind of weighted jet veto over phase space regions based on the kinematics of the hard process. Moreover, by choosing different weights the moments can be

adjusted such that they avoid introducing a fixed scale below the factorization scale of the hard process. At the expense of the background rejection efficiency they can be tuned to introduce smaller theory uncertainties. This will allow the ATLAS and CMS experiments to optimize their Higgs analyses including theory uncertainties and significantly improve the case for a luminosity upgrade based on Higgs couplings measurements.

In this paper we will attempt to answer three questions based on the weak boson fusion analysis with a Higgs decay to photons. This includes a study of the signal process, the Higgs background from gluon fusion, and the continuum production of a photon pair with jets:

1. in Section 2.2.3 we will apply Fox–Wolfram moments to the kinematics of the two tagging jets only. Based on a multivariate analysis we will estimate how much these additional observables can improve the current ATLAS results at 8 TeV collider energy.

2. in Section 2.2.4 we will compare the performance of a set of Fox–Wolfram moments with a specific (unit) weight in comparison to the usual central jet veto for the 13 TeV run. Moreover, a multivariate analysis of Fox-Wolfram moments allows us to define a ROC curve with a free choice of operating points.

3. in Section 2.2.5 we will introduce a new weight in the Fox–Wolfram moments. It avoids introducing a physical momentum scale for the jet veto which lies below the factorization scale.

Obviously, our conclusions are immediately applicable to ongoing and future LHC analyses. Fox–Wolfram moments have been tested in a few ATLAS and CMS analyses, so it should be a simple task to also include them in Higgs analyses.

2.2.2 Setting the Stage

The analysis presented in this paper will give an estimate of the impact which Fox–Wolfram moments computed from jets can have on current and future LHC Higgs analyses. Fox–Wolfram moments are one way to systematically evaluate angular correlations between jets in terms of spherical harmonics. While such approaches are standard for example in cosmology, they are largely missing in LHC physics. We will summarize their main features below. For a more detailed account of the WBF-specific properties we refer to an earlier paper [159].

To allow for significant correlations between different moments we employ multivariate methods. Our analysis will largely be based on boosted decision trees (BDTs), which we will also briefly introduce below. Part of the analysis we cross–check with a neural net to make sure our findings are independent of the MVA method used.

Fox–Wolfram moments

Most analyses of QCD jets at the LHC are based on an *ad-hoc* selection of angular correlation variables, which have been shown to separate signals from backgrounds. For analyses where each one–dimensional or two–dimensional distribution is carefully understood in terms of the underlying physics and then tuned to the best cut value, this approach is natural and appropriate. For multivariate analyses, where events are classified in terms of a more generic set of kinematic observables, the choice of observables should be more systematic.

For angular correlations, we know how to generally describe underlying objects, in our case jets, in terms of spherical harmonics. Obviously, Fox–Wolfram moments do not have to be based on jets. They are closely related to event shapes [161, 162], and for example at LEP they were based on calorimeter information. At the LHC, particle flow objects or topoclusters might eventually turn out more useful. In this analysis we use jets to avoid additional experimental or theoretical complications, for example due to pile-up or underlying event.

Fox–Wolfram moments are constructed by summing jet–jet correlations over all $2\ell+1$ directions, including an unspecified weight function W_i^x [158]

$$H_\ell^x = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} \left| \sum_{i=1}^{N} W_i^x \, Y_\ell^m(\Omega_i) \right|^2 \quad . \tag{2.67}$$

The index i sums over all final state jets defined by appropriate acceptance and selection criteria. The general coordinates of the spherical harmonics $Y_\ell^m(\theta,\phi)$ we replace by a reference angle Ω. The moments can be rewritten as

$$H_\ell^x = \sum_{i,j=1}^{N} W_{ij}^x \, P_\ell(\cos\Omega_{ij}) \qquad \text{with} \quad W_{ij}^x = W_i^x W_j^x \; . \tag{2.68}$$

The angle Ω_{ij} is the total angle between two jets, which implies that it cannot be boost invariant. The weight function W_{ij}^x can be chosen freely. In Sections 2.2.3 and 2.2.4 we will use transverse–momentum and unit weights [159]:

$$W_{ij}^T = \frac{p_{Ti}\, p_{Tj}}{(\sum p_{Ti})^2} \qquad W_{ij}^U = \frac{1}{N^2} \; . \tag{2.69}$$

The advantage of the transverse–momentum weight is that soft and collinear jets with their limited amount of information about the hard process are automatically suppressed. The resulting analysis becomes stable with respect to the parton shower and QCD jet radiation. For tagging jets without an actual collinear divergence the transverse momentum weight should be appropriate.

Whenever we are interested in the color structure of the event, this jet radiation will carry the crucial information. For studies of central jet radiation we therefore expect the unit weight to be the most promising.

In analogy to a jet veto, Fox–Wolfram moments with unit weight introduce an energy or momentum scale, above which we include jets in the moments. Because of the unit weight there does not exist a smooth transition regime; requiring any Fox–Wolfram moment of such additional jets to be different from zero corresponds to a step function in counting the number of jets. Because the new momentum scale usually resides below the factorization scale of the hard event, fixed–order precision predictions are not applicable, and a dedicated resummation is a theoretical challenge [103, 136, 154–157]. In Section 2.2.5 we introduce the *matched weight*

$$W_{ij}^M = \frac{(p_{Ti} - p_T^{\min})\,(p_{Tj} - p_T^{\min})}{\left(\sum p_{Ti} - p_T^{\min}\right)^2} \tag{2.70}$$

in order to reduce the theoretical uncertainty in comparing measured cross sections to QCD predictions. This new weight avoids introducing a new hard scale and will be less dominated by the momentum scale $p_{Tj}^{\min} = 20$ GeV, above which jets contribute to the Fox–Wolfram moments.

Event generation

While the description of the tagging jets in weak boson fusion is straightforward, the continuum background with its QCD jet activity is more tricky. Moreover, the correct description of the QCD activity in the Higgs signal requires a careful treatment of the color structure of the hard process. Throughout this analysis we use Sherpa [47] with Ckkw merging [45]. For the weak boson fusion signal we generate samples including up to three hard jets, including the tagging jets. Gluon fusion Higgs production we simulate with up to three hard jets. For the QCD background we include di–photon production plus up to two hard jets. For jet clustering we rely on the anti-k_T algorithm as in Fastjet [38] with $R = 0.4$.

The assumed Higgs mass value is 126 GeV. Our cuts are dominated by the detector acceptance and jet–photon separation,

$$p_{T\gamma} > 14 \;\; \text{GeV} \qquad R_{\gamma j} > 0.3 \qquad m_{\gamma\gamma} > 80 \;\; \text{GeV} \; . \tag{2.71}$$

After those cuts we are left with a weak–boson–fusion signal cross section times branching ratio of 5.2 fb at 8 TeV collider energy and 9.24 fb at 13 TeV collider energy. To allow for an efficient generation of background events we do not require a mass window for the two photons in the background generation. Later in the analysis we add an $m_{\gamma\gamma}$ window of ± 10 GeV around the Higgs mass. For a proper Higgs analysis we should require an $m_{\gamma\gamma}$ window of 1-2 GeV around the measured Higgs mass. However, with this condition the event generation for the background becomes highly inefficient. Because our analysis does not intend to predict the actual signal and background cross sections and instead focuses on the improvement over the established experimental analysis [164], the loose cuts of Eq.(2.71) allow for a much more efficient event generation and will not affect our conclusions.

Boosted decision trees

Any multivariate analysis is based on some kind of mapping of a set of observables onto a single–valued quantity, the classifier response. Based on this classifier response we define a classification rule to separate signal and background events. Training the multivariate analysis on a set of simulated events aims to determine the best classification rule for a given signal and background. The optimal classification rule has to be determined by some measure, for example the signal efficiency, the statistical significance, or the signal–to–background ratio. Independent of this optimization, we can quantify the performance of any classification rule in terms of the signal efficiency and the background mis-identification probability. In this two–dimensional plane we can describe cuts on the same response parameter as a receiver operating characteristics (ROC) curve. Given such a ROC curve we are free to choose one or more operating points. In line with the ATLAS di-photon analysis we use a fixed 40% signal efficiency ϵ_S after acceptance with a variable background rejection $1-\epsilon_B$ as the standard working point. In Section 2.2.4, we quote the main results of our BDT analysis for the best possible significance $S/\sqrt{S+B}$ given the set of kinematic observables and Fox–Wolfram moments.

Decision tree algorithms — as they are utilized in high energy physics applications — are based on a set of kinematic variables, intended to separate signal and background events. In the first step they choose the 'root node' variable, i.e. the variable with the best separation between signal and background. There exist several types of separation which we can choose from in TMVA [163]. We use the cross entropy

$$C_E = -\frac{S}{S+B}\log_2\frac{S}{S+B} - \frac{B}{S+B}\log_2\frac{B}{S+B}\,, \tag{2.72}$$

where S and B are the numbers of signal and background events in a particular subset of events. This measure is the closest to the original definition of information entropy [165]. After choosing the root node, the subsequent nodes are ordered by their separation at some threshold value.

For the complete decision tree the events are classified as signal–like or background–like by some measure. In the training set we know how good the tree is at classifying the events. Our training set include 100000 events for each signal and background channel. In the next step the algorithm corrects for mistakes through a reweighting procedure, builds another decision tree, tests its performance, and repeats for some user–defined number of iterations. For this 'boosting' procedure we mainly use the adaptive boost algorithm implemented in TMVA [163]. The final classification rule for signal versus background events we then apply to an independent event sample, again including 100000 events per signal and background process. To prevent over–training we limit our forest to 400 trees, and the individual trees to three layers.

Because correlations between the different Fox–Wolfram moments are a key issue of our systematic approach to kinematic input variables, we carefully test two different boosting algorithms (adaptive and gradient boost [163]) as well as different multivariate analysis methods. *Per se*, boosted decision trees are not particularly well suited for studying strongly correlated variables. The reason is that trees are built out of the individual variables. Two strongly correlated variables are best mapped through individual fine binnings in each of them, so a careful mapping of correlations will eventually lead to statistical limitations and a possible training on statistical fluctuations. Therefore, we compare BDT results to results using a multi–layer perceptron (MLP) neural network whenever an independent test appears sensible. We utilize a MLP neural network with a single hidden layer containing $N+5$ neurons, where N is the number of training variables.

2.2.3 Tagging jet correlation

In this first analysis we are going to use Fox–Wolfram moments to systematically test the completeness of the tagging jet correlations included by ATLAS. Because we directly refer to the current ATLAS result we use a collider energy of 8 TeV for the most recent LHC run. The two p_T-ordered tagging jets have to fulfill either of the two conditions

$$\begin{aligned} p_{Tj} &> 25 \text{ GeV} \quad \text{for} \quad |y_j| < 2.4 \\ p_{Tj} &> 30 \text{ GeV} \quad \text{for} \quad 2.4 \leq |y_j| < 4.5\,. \end{aligned} \tag{2.73}$$

	BDT			MLP		
$\epsilon_S = 0.4$	$1-\epsilon_B$	$\frac{S}{\sqrt{S+B}}$	$\frac{S}{B}$	$1-\epsilon_B$	$\frac{S}{\sqrt{S+B}}$	$\frac{S}{B}$
ATLAS default Eq.(2.75)	0.887	1.50	0.76	0.888	1.50	0.78
$H_1^{T,\phi} \to H_4^{T,\phi}$, $H_1^{U,\phi} \to H_4^{U,\phi}$	0.952	1.65	1.54	0.953	1.65	1.55
$H_1^{T,\phi}$, $H_3^{T,\phi}$, $H_1^{U,\phi}$, $H_3^{U,\phi}$	0.952	1.66	1.56	0.952	1.65	1.54
$H_1^{T,\phi}$, $H_2^{T,\phi}$, $H_2^{U,\phi}$, $H_2^{U,\phi}$	0.953	1.65	1.47	0.953	1.65	1.55
$H_1^{T,\phi}$, $H_1^{U,\phi}$	0.953	1.65	1.43	0.952	1.65	1.46
$H_1^{T,\phi}$	0.950	1.63	1.45	0.950	1.63	1.44
$H_1^{U,\phi}$	0.952	1.65	1.40	0.952	1.65	1.44
$\cos\Delta\phi_{12}$, W_{12}^T	0.952	1.65	1.53	0.952	1.65	1.50
$\cos\Delta\phi_{12}$	0.952	1.65	1.42	0.952	1.65	1.44

Table 2.6: BDT and MLP results including azimuthal–angle Fox–Wolfram moments based on the two tagging jets only after Eq.(2.74). The background rejection is given for 40% signal efficiency. The value for $S/\sqrt{S+B}$ we compute for an integrated luminosity of 30 fb^{-1}. All sets of variables subsequent to the first row contain the default variables as well.

These two tagging jets must also pass

$$|\Delta y_{j_1 j_2}| \geq 2 \quad \text{and} \quad m_{j_1 j_2} > 150 \text{ GeV} . \tag{2.74}$$

These cuts correspond to the variables used in the multivariate di-photon Higgs analysis by ATLAS [164],

$$\{m_{j_1 j_2}\,,\, y_{j_1}\,,\, y_{j_2}\,,\, \Delta y_{j_1 j_2}\} \qquad \text{(ATLAS default)}. \tag{2.75}$$

The angular correlations between the tagging jets in weak–boson–fusion Higgs production is known to reflect the tensor structure of the WWH vertex [160]. In this application the collinearity of the two tagging jets plays an important role, with the effect that the azimuthal angle between the tagging jet is a more sensitive probe than the opening angle between them. For the Fox–Wolfram moments this means that the definition in terms of the opening angle Ω_{ij} is not optimally suited. For the tagging jet analysis we therefore replace the opening angle in the Legendre polynomials by the azimuthal angle $\Delta\phi_{ij}$ between the two tagging jets,

$$H_\ell^{x,\phi} = \sum_{i,j=1}^{2} W_{ij}^x \, P_\ell(\cos\Delta\phi_{ij}) \, . \tag{2.76}$$

For a systematic study of the usefulness of the tagging jet correlations we perform a multi-variate analysis of the Fox–Wolfram moments introduced in Section 2.2.2. Because the moments are based on spherical harmonics they form a basis and include all available information, given the weight W_{ij}^x we use in their definition.

We show some sample BDT and MLP results based on the azimuthal moments in Table 2.6. The full set of moments for each weight function by definition includes all available information for the corresponding weights. First, we see that including a large set of Fox–Wolfram moments gives a significant improvement of the current ATLAS set of observables, defined in Eq.(2.75). Both multivariate analyses using the first four moments with unit weight as well as with transverse–momentum weight reduces the remaining fraction of background events by a factor two. From the TMVA output we have checked that these eight moments dominate the distinctive power of the analysis.

Obviously, the next question is which of the Fox–Wolfram moments contribute most to this improvement. From the earlier analysis [159] we know that lower moments will dominate in the tagging jet analysis, and that only odd moments can distinguish between forward–backward and forward–forward tagging jets. Individually, we find that the six best individual moments are (in

order) $H_1^{U,\phi}$, $H_1^{T,\phi}$, $H_3^{U,\phi}$, $H_3^{T,\phi}$, $H_2^{U,\phi}$, and $H_2^{T,\phi}$.[42] The moments with unit weight are slightly more powerful than the transverse–momentum weight. The most striking feature is that for the tagging jet the higher moments play hardly any role in improving the analysis.

As a matter of fact, the single moment $H_1^{U,\phi}$ is, within uncertainties due to the training procedure, almost as powerful as the set of the first 20 moments, both with unit and transverse–momentum weight. Given that the corresponding Legendre polynomial is $P_1(\cos\Delta\phi_{ij}) = \cos\Delta\phi_{ij}$ we can further simplify the analysis by separating the transverse–momentum weight from the azimuthal angle. Compared to the ATLAS default variables, adding the azimuthal angle between the tagging jets, $\Delta\phi_{ij}$, almost doubles the signal–to–background ratio. Systematically including the Fox–Wolfram moments increases the signal–to–background ratio additionally by 8%. This result persists between the two multivariate methods and we conclude that our improvement is truly due to the nature of the moments and not to some advantageous choice of methods and/or parameters for our multivariate analyses.

Following the tagging jet analysis in this section we extend the default set of tagging jet cuts Eq.(2.75) for the remainder of this paper to include

$$\{m_{j_1j_2}\,,\, y_{j_1}\,,\, y_{j_2}\,,\, \Delta y_{j_1j_2},\, \Delta\phi_{j_1j_2}\} \qquad \text{(WBF default)}. \tag{2.77}$$

It could be argued that adding the azimuthal angle to the list of kinematic variables employed in the background rejection will make the analysis result less applicable to modified Higgs–like signal hypotheses. Indeed, the azimuthal angle between the tagging jets is the key observable in the spin-0 CP analysis of the Higgs resonance [160]. On the other hand, the same is true for the rapidity difference Δy_{12} when it comes to spin-2 alternatives [160].

2.2.4 Replacing a jet veto

The key physics question we will answer in this Section is to what degree we can use information on additional (central) jet radiation to enhance the tagging jet analysis described in the previous Section 2.2.3. Because a detailed analysis of the jet activity has not been performed in the recent LHC runs, we assume a collider energy of 13 TeV in this section. The physics of the additional jets can be easily described: for the signal events the emission of additional central jets is suppressed by the color structure of the process. This means that the number of jets in weak boson fusion will in general follow the staircase pattern predicted for inclusive processes at the LHC [154]. In contrast, gluon–fusion Higgs production or di-photon production will show this staircase pattern only in the absence of tagging jet cuts. Once we require two hard jets with a large invariant mass we induce large logarithms, which leads to a Poisson pattern in the number of jets [154]. The key feature of this Poisson distribution is a significantly enhanced probability of radiating a central jet.

Throughout our analysis we require two tagging jets with the generic acceptance cuts

$$p_{Tj} > 20 \text{ GeV} \qquad |y_j| < 4.5 \tag{2.78}$$

$$|\Delta y_{j_1j_2}| > 2 \qquad m_{j_1j_2} > 150 \text{ GeV}\,. \tag{2.79}$$

Correspondingly, we generate signal and background events using Sherpa [47] with CKKW [45] jet merging with two or three hard jets from the matrix element. Throughout this Section we assume a collider energy of 13 TeV. In addition to the general photon cuts of Eq.(2.71) we require $m_{\gamma\gamma} = 126 \pm 10$ GeV. The cuts of Eq. (2.78) lead to cross sections of 6.5 fb for the weak–boson–fusion signal, 4.5 fb for gluon–fusion Higgs production, and 2050 fb for the continuum background. As mentioned above, the signal–to–background ratio can be improved through additional cuts, such as tightening the $m_{\gamma\gamma}$ requirement. However, this makes it harder to reliably simulate the background. In the following we will assume that additional cuts on the Higgs decay products are orthogonal to the additional jet kinematics.

Because the selection criterion of the two tagging jets has a significant impact on the amount of Poisson enhancement of the additional jet production we use two selection criteria for the tagging jets:

1. p_T-selection: of all jets fulfilling Eqs.(2.78) and (2.79) the two hardest are the tagging jets. The mild cuts of Eq.(2.79) leave 3.36 fb for the signal, 1.04 fb for gluon–fusion Higgs production, and 509 fb for the continuum background.

[42] Given that Tmva gives an ordered list of the most relevant observables, it is not clear to one of the authors (TP) why this very interesting information is never shown in experimental publications.

	Δy-selection			p_T-selection		
	WBF	GF	$\gamma\gamma$	WBF	GF	$\gamma\gamma$
generated [fb]	6.5	4.5	2050	6.5	4.5	2050
$\Delta y_{j_1 j_2} > 4.4$	$\times 0.33$	$\times 0.15$	$\times 0.11$	$\times 0.27$	$\times 0.056$	$\times 0.055$
$y_{j1} y_{j2} < 0.0$	$\times 1.00$	$\times 1.00$	$\times 1.00$	$\times 1.00$	$\times 1.00$	$\times 1.00$
$m_{j_1 j_2} > 600$ GeV	$\times 0.72$	$\times 0.55$	$\times 0.46$	$\times 0.77$	$\times 0.61$	$\times 0.47$
cut level [fb]	1.52	0.37	107	1.36	0.15	52.9
central jet veto	$\times 0.75$	$\times 0.15$	$\times 0.22$	$\times 0.91$	$\times 0.45$	$\times 0.52$
veto level [fb]	1.14	0.056	24.0	1.24	0.068	27.7

Table 2.7: Cut flow for the standard weak–boson–fusion analysis with a central jet veto for an LHC energy of 13 TeV.

	Δy-selection				p_T-selection			
	ϵ_S	$1-\epsilon_B$	$\frac{S}{\sqrt{S+B}}$	$\frac{S}{B}$	ϵ_S	$1-\epsilon_B$	$\frac{S}{\sqrt{S+B}}$	$\frac{S}{B}$
acceptance cuts Eqs.(2.78) and (2.79)	1	0	0.76	0.005	1	0	0.81	0.007
veto–level cuts Eq.(2.80)	0.402	0.854	0.80	0.014	0.405	0.996	1.01	0.026
jet veto	0.302	0.967	1.24	0.047	0.369	0.945	1.26	0.045
BDT: WBF default with Eq.(2.79)	0.400	0.862	0.79	0.014	0.400	0.904	1.04	0.027
	0.634	0.674	0.84	0.010	0.414	0.897	1.04	0.027
BDT: WBF default plus FWM with Eq.(2.79)	0.400	0.952	1.34	0.041	0.400	0.944	1.35	0.047
	0.232	0.986	1.42	0.083	0.302	0.972	1.43	0.071

Table 2.8: S/B and $S/\sqrt{S+B}$ compared to classical cut and jet veto strategy for the Δy and p_T-selections of the tagging jets. The value for $S/\sqrt{S+B}$ we compute for an integrated luminosity of 30 fb^{-1}. The BDT analysis includes a set of Fox–Wolfram moments with unit weight, Eq.(2.81). We quote two working points at 40% signal efficiency and optimized for $S/\sqrt{S+B}$.

2. Δy-selection: of all jets fulfilling Eq.(2.78) and (2.79) the two most forward and backward are the tagging jets, maximizing $\Delta y_{j_1 j_2}$. After Eq.(2.79) the remaining rates are 3.78 fb for the signal, 1.71 fb for gluon–fusion Higgs production, and 736.2 fb for the non-Higgs background.

While the p_T-selection is standard in most weak–boson–fusion analyses, it will turn out that the Δy-selection is more efficient in generating a large Poisson enhancement for central jet emission in the background processes. On the other hand, in particular for the 13 TeV run we have to see if pile-up makes one of the two selections appear experimentally superior.

The standard approach to including the additional jet activity in the weak–boson–fusion Higgs analysis is a central jet veto [153, 156]. To generate a sufficiently strong Poisson pattern in the number of jets we demand

$$|\Delta y_{j_1 j_2}| > 4.4 \qquad y_{j1} \cdot y_{j2} < 0 \qquad m_{j_1 j_2} > 600 \text{ GeV} \,. \tag{2.80}$$

In Table 2.7 we show the cut flow of the signal and background rates for each step in Eq.(2.80). Finally, we include a central jet veto which does not allow for jets above $p_T = 20$ GeV in between the two tagging jets. While the two tagging jet selections show significant differences in the intermediate steps, after the veto the numbers of signal and background events are comparable. The survival rates for the central jet veto are in agreement with the literature [149, 156].

In the first three rows of Table 2.8 we show different statistical measures after the acceptance cuts of Eqs.(2.78) and (2.79), the veto–level cuts of Eq.(2.80), and after the central jet veto. The background is composed of gluon–fusion Higgs production and continuum di-photon production. We again see that the significance $S/\sqrt{S+B}$ and the signal–to–background ratio are comparable for the Δy-selection and the p_T-selection of the tagging jets. However, this is only true after the jet veto. After only the hard cuts of Eq.(2.80) the p_T-selection is significantly more promising. As alluded to above, the jet veto benefits from the stronger Poisson enhancement from the Δy-selection, leaving the final results essentially identical.

In the next step, we use the default WBF observables of Eq.(2.77) and optimize them in a multivariate BDT analysis as described in Section 2.2.2. The corresponding ROC curve we show in Figure 2.28. As in Table 2.8 the efficiencies are defined with respect to the full set of acceptance cuts from Eqs.(2.78) and (2.79). In the table we quote two points from this curve. First, we show the usual working point with a signal efficiency of 40%. Second, we show the working point

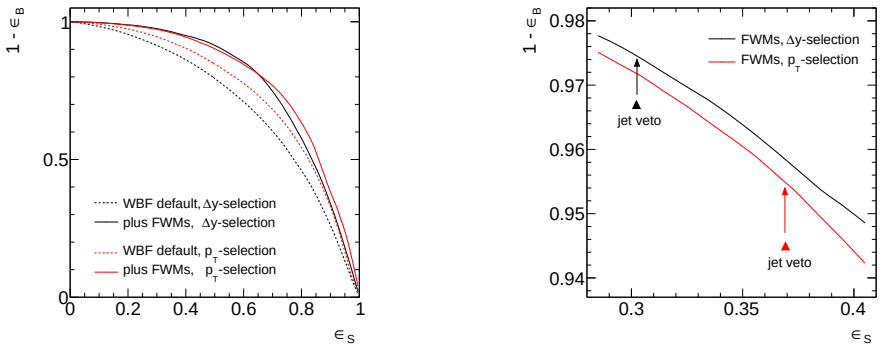

Figure 2.28: ROC curve for Δy- (black) and p_T-selection (red) of the tagging jets. Left: We compare the WBF default observables (dashed) of Eq.(2.77) to an additional set of Fox–Wolfram moments (solid). Right: We show how using Fox–Wolfram moments compare to a central jet veto.

with the best result for $S/\sqrt{S+B}$. Optimizing for the best result of S/B does not give a well defined solution. As expected, the ROC curve indicates working points for the entire range of signal efficiencies $\epsilon_S = 0...1$.

The question we attempt to answer in this section is if we can use the available information on the additional jet activity in reducing the two backgrounds more efficiently than with a central jet veto. The baseline for this comparison is the corresponding row in Table 2.8. As described in Section 2.2.2 we rely on a large set of Fox–Wolfram moments forming a basis for the angular correlations given a weight W_{ij}^x. Unlike for the tagging jet kinematics we now do not constrain our system to the transverse plane, which means we use the original definition of the moments in Eq.(2.68) with the opening angle Ω_{ij}. On the other hand, we already know what the benefit of including the moments of the tagging jets are: according to Section 2.2.3 most of the information is included once we add the azimuthal angle between the tagging jets, $\Delta\phi_{j_1 j_2}$, to the standard set of observables given in Eq.(2.77). Therefore, we limit the analysis of the additional jet activity to all jet–jet correlations *with the exception of the two tagging jets*. Moreover, we can expect the unit weight to give the best sensitivity to the relatively soft additional jet activity, so we use

$$H_\ell^U = \frac{1}{N^2} \sum_{(i,j)\neq(1,2)} P_\ell(\cos\Omega_{ij}) \; . \tag{2.81}$$

For both of the tagging jet selections we only include jets which fall between the two tagging jets, in complete analogy of a central jet veto. For exactly two tagging jets and no additional jet radiation this implies $H_\ell^U = 0$ for all values of ℓ.

In Table 2.8 we show the result of a combined BDT analysis of the observable of Eq.(2.77) and the set of Fox–Wolfram moments. Again, we quote two operating points, one of them for a fixed signal efficiency of 40% and one optimized for the best value of $S/\sqrt{S+B}$. In addition, we show results for both, the Δy-selection and the p_T-selection of the tagging jets. A generic problem for any BDT analysis is that for limited statistics of the training sample it can only include a limited number of observables. On the other hand, the BDT first determines the most powerful observables, so we only include the five best Fox–Wolfram moments in our analysis. We have checked that adding more moments will not improve the result beyond numerical accuracy. For the Δy-selection the five leading moments with unit weight are H_2^U, H_4^U, H_{18}^U, H_{19}^U, and H_{17}^U. For the p_T-selection the most powerful moments are H_2^U, H_{19}^U, H_{17}^U, H_{20}^U, and H_{15}^U. However, for the p_T-selection the most powerful variable in the BDT is $\Delta y_{j_1 j_2}$. For the Δy-selection this observable is maximized by construction.

The ROC curves in Figure 2.28 shows a clear improvement of the complete multivariate analysis including the Fox–Wolfram moments as compared to the kinematic variables of Eq.(2.77) only. For a fixed moderate signal efficiency of 40% adding information on the jets decreases the probability of a background mis-identification by a factor of 2.9 for the Δy-selection and a factor of 1.7 for the p_T-selection. The improvement relative to the jet veto we show in the right panel, zooming into typical signal efficiencies around 35% relative to the acceptance cuts of Eq.(2.79). For the jet veto

working point of the Δy-selection with fixed signal efficiency of 30.2% we see that the background misidentification is reduced by 30%. For the p_T-selection with fixed signal efficiency of 36.9% we find an improvement by 20%.

2.2.5 Avoiding new scales

The unit weights in the definition of the Fox–Wolfram moments used in the previous Section 2.2.4 share a disadvantage with a jet veto when it comes to predicting them from theory: they introduce an additional physical momentum scale in the process which is below the hard scale of the Higgs production process. In Eq.(2.78) a transverse momentum scale of 20 GeV is introduced as the minimum transverse momentum of the jets contributing to the moments. Collinear factorization as the basis of defining the parton densities in perturbative field theory does not allow us to consider the details of jets which are softer than the hard scale chosen to describe the process in perturbative QCD. All measurements which are to be compared to fixed–order perturbative QCD predictions have to be jet–inclusive for transverse momenta below the factorization scale. If we introduce an additional energy scale below the factorization scale this implies that we introduce a possibly large logarithm which needs to be resummed [103, 136, 157].

Introducing a weight which smoothly interpolates between the jet counting scale $p_{Tj}^{\min} = 20$ GeV and the hard scale of the process according to Eq.(2.70) should alleviate this tension, suggesting to repeat the same analysis as shown in Section 2.2.4 with the Fox–Wolfram moments

$$H_\ell^M = \sum_{(i,j)\neq(1,2)} \frac{(p_{Ti} - p_T^{\min})\,(p_{Tj} - p_T^{\min})}{\left(\sum p_{Ti} - p_T^{\min}\right)^2} P_\ell(\cos\Omega_{ij}) \; . \tag{2.82}$$

The slow turn–on of these matched Fox–Wolfram moments should soften the impact of the additional energy scale. While we cannot offer an estimate of the improvement in the perturbative QCD treatment, we expect that the matched weights are less sensitive to large collinear logarithms generated by the apparent violation of collinear factorization. While a full theoretical analysis of the benefits of these new, matched moments is beyond the scope of this paper, we will estimate how their experimental benefits compare to the unit weight moments shown above.

In Table 2.9 we extend the original Table 2.8, including the same BDT analysis now based on matched Fox–Wolfram moments. For the standard working point with 40% signal efficiency we see that the background rejection from the matched moments is essentially identical to the unit weight moments. The main difference is the order of the most relevant set of moments, which now is $H_1^M, H_2^M, H_3^M, H_4^M, H_6^M$ for the Δy-selection and $H_1^M, H_3^M, H_6^M, H_2^M, H_4^M$ for the p_T-selection. Similarly, the working point optimized for $S/\sqrt{S+B}$ is only slightly shifted. In Figure 2.29 we compare the ROC curves for the jet radiation study based on the two Fox–Wolfram moment weights. For signal efficiencies between 25% and 40% the unit weight is slightly superior, but most likely this slight advantage will be compensated once we include theory uncertainties from QCD predictions.

2.2.6 Outlook

Weak boson fusion analyses of Higgs production at the LHC are key ingredients to Higgs couplings and Higgs property analyses in the upcoming LHC run. They allow for an efficient

	Δy-selection				p_T-selection			
	ϵ_S	$1-\epsilon_B$	$\frac{S}{\sqrt{S+B}}$	$\frac{S}{B}$	ϵ_S	$1-\epsilon_B$	$\frac{S}{\sqrt{S+B}}$	$\frac{S}{B}$
jet veto Eq.(2.78) to (2.80)	0.302	0.967	1.24	0.047	0.369	0.945	1.26	0.045
BDT: WBF default plus unit–weight FWM	0.400	0.952	1.34	0.041	0.400	0.944	1.35	0.047
	0.232	0.986	1.42	0.083	0.302	0.972	1.43	0.071
BDT: WBF default plus matched–weight FWM	0.400	0.949	1.32	0.040	0.400	0.942	1.32	0.045
	0.240	0.985	1.43	0.081	0.256	0.979	1.40	0.082

Table 2.9: S/B and $S/\sqrt{S+B}$ compared to jet veto strategy for the Δy and p_T-selections of the tagging jets. The value for $S/\sqrt{S+B}$ we compute for an integrated luminosity of 30 fb^{-1}. Extending Table 2.8 the BDT analysis now includes a set of Fox–Wolfram moments with matched weight, Eq.(2.82). As BDT results we quote the working point at 40% signal efficiency and the best point for $S/\sqrt{S+B}$.

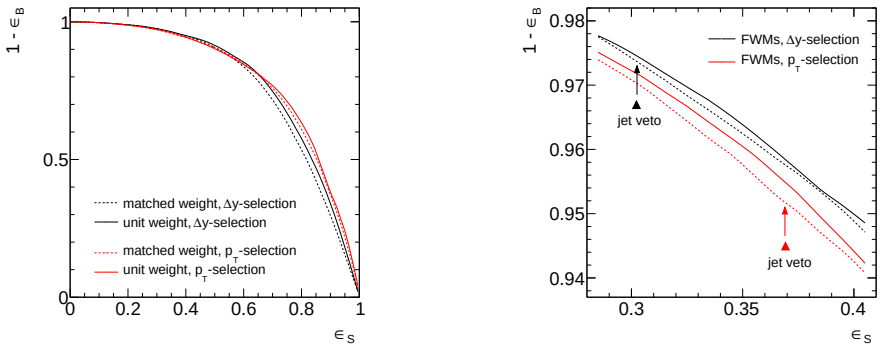

Figure 2.29: ROC curve for Δy- (black) and p_T-selection (red) of the tagging jets. Left: We compare the choice of unit weight in the FWMs (solid) to a p_T matched weight (dashed). Right: We show how using Fox–Wolfram moments with different weights compares to a central jet veto.

background rejection based on two tagging jets and an additional central jet veto. The question is, how we can make optimal use of the jet properties for example to improve the signal–to–background ratio or the signal significance. In our detailed analysis we come to three conclusions:

1. For the two tagging jets we rely on a set of low-ℓ moments with a transverse momentum weight and azimuthal angle separation. Most of the improvement as compared to the standard ATLAS analysis can be traced back to the missing azimuthal angle between the tagging jets. In addition, the signal–to–background ratio can be increased by 8% by including a set of Fox–Wolfram moments.

2. The additional jets can be studied using a wide range of moments with a unit weight and full angular separation. It should be compared to a jet veto and delivers a significantly better performance. The tagging jet selection with maximum rapidity distance is better suited to distinguish the signal from the continuum background then the transverse momentum selection. For both cases we computed a full ROC curve, allowing for optimized working points depending on the details of the analysis.

3. To reduce theory uncertainties from QCD predictions we can introduce a softer, matched weight in the Fox–Wolfram moments. It turns out that the analysis of jet radiation is almost as promising as for the unit weights, but with a much improved theoretical behavior.

We conclude that tagging jet criteria as well as the jet veto as analysis tools for Higgs analyses in weak boson fusion can be improved by a systematic study of the multi–jet system based on Fox–Wolfram moments. The improvement is significant, both for the Δy-selection and the p_T-selection of the tagging jets. The Fox–Wolfram moment analysis can be adapted to individual analyses by choosing appropriate working points in the corresponding ROC curves.

2.2.7 Discriminative power

In the sections above we study how FWMs together with a BDT analysis enhance Higgs WBF searches. We find two main improvements. First we can boost the tagging jet analysis, because standard analyses miss an important observable, $\Delta\Phi_{jj}$. Second we find that we can replace a jet-veto by using the FWMs. This results in a significant enhancement in S/B. In this sub section we address the question of: How much enhancement in S/B comes from which particular source, e.g. choice of TMVA method, set of observables, etc. The results presented here originate from fruitful discussion with Jamie Tattersall. For simplicity we only consider the so called Δy-selection and the unit weight prescription, see [33]. Furthermore, if not stated explicitly, we always train on the WBF default set of observables, defined by

$$m_{j_1 j_2}\,,\; y_{j_1}\,,\; y_{j_2}\,,\; \Delta y_{j_1 j_2},\; \Delta\phi_{j_1 j_2} \qquad \text{(WBF default)}. \tag{2.83}$$

We denote the identified tagging jets with 1 and 2. To be in accordance with [33] and the ATLAS default we require the following generic acceptance cuts

$$p_{T,j} > 20 \;\text{ GeV}, \qquad |y_j| < 4.5, \tag{2.84}$$

$$|\Delta y_{j_1 j_2}| > 2, \qquad m_{j_1 j_2} > 150 \;\text{ GeV}\,. \tag{2.85}$$

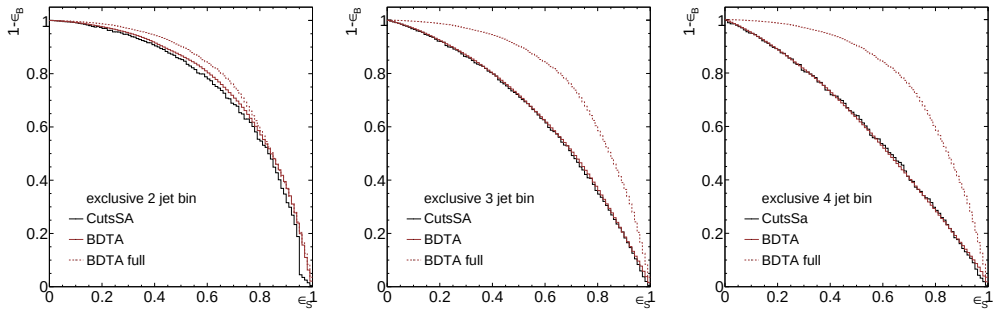

Figure 2.30: Comparison of boxed cuts (black) vs. BDT (red) for exclusive jet bins ($n_{\text{jets}} = 2, 3, 4$). The dashed red line represents the full n_{jets} BDT analysis for comparison.

These cuts lead to the following cross sections: $\sigma_{\text{WBF}} = 3.78$ fb, $\sigma_{\text{GF}} = 1.71$ fb, $\sigma_{\gamma\gamma} = 736.2$ fb.

Let us now deconstruct our previous analysis. We know that the construction of the FWMs is sensitive to the number of jets. It is easy to check that using n_{jets} in the BDT analysis instead of the FWM gives a similar (but still worse) ROC curve. Compare the red and green lines in Fig. 2.31. If we use n_{jets} in a BDT analysis TMVA tells us that it is the most discriminating observable in the tree. The next step is to quantify which of the enhancement is due to the inclusion of the number of jets respectively the FWM into our strategy and which part is from the BDT. It is a well established strategy to run different Higgs searches in different jet bins. Here, however, we focus on a combined search for several bins. We understand combined in the sense that we do not optimize separately in each exclusive jet channel, but combine ϵ_S and ϵ_B of the different channels. For each exclusive jet bin we run a simple optimized boxed cut analysis. In addition we run a BDT analysis in the same jet bin. We find that the cut based analysis is nearly as good as the BDT, see Fig 2.30. The question is: to what degree can we translate this to the full analysis, where the jet spectrum is just another variable? To compare them we have to melt the different ROC curves of BDT-per-jet and cuts-per-jet into one object each. We do this via the following description:

$$\begin{aligned} \epsilon_{\text{S}} &= \sum_i \epsilon_{\text{S}}^i \frac{\sigma_{\text{S},i}}{\sigma_{\text{S,tot}}}, \\ \epsilon_{\text{B}} &= \sum_i \epsilon_{\text{B}}^i \frac{\sigma_{\text{S},i}}{\sigma_{\text{B,tot}}}. \end{aligned} \tag{2.86}$$

From our MC study we use the fractions

$$\begin{aligned} \frac{\sigma_{\text{S},2}}{\sigma_{\text{S,tot}}} &= 0.71, & \frac{\sigma_{\text{S},3}}{\sigma_{\text{S,tot}}} &= 0.19, & \frac{\sigma_{\text{S},4}}{\sigma_{\text{S,tot}}} &= 0.07, \\ \frac{\sigma_{\text{B,GF},2}}{\sigma_{\text{B,GF,tot}}} &= 0.27, & \frac{\sigma_{\text{B,GF},3}}{\sigma_{\text{B,GF,tot}}} &= 0.30, & \frac{\sigma_{\text{B,GF},4}}{\sigma_{\text{B,GF,tot}}} &= 0.20, \\ \frac{\sigma_{\text{B},\gamma\gamma,2}}{\sigma_{\text{B,tot},\gamma\gamma}} &= 0.35, & \frac{\sigma_{\text{B},\gamma\gamma,3}}{\sigma_{\text{B,tot},\gamma\gamma}} &= 0.34, & \frac{\sigma_{\text{B},\gamma\gamma,4}}{\sigma_{\text{B,tot},\gamma\gamma}} &= 0.19. \end{aligned}$$

We plot all possible combinations into one graph. This gives us not a ROC curve but an ROC area. The boundary of this area is our combined ROC curve, see Fig. 2.31. We observe that n_{jets} (red) and FWM (green) agree with a very slight advantage for the FWM. Furthermore, the deconstruction in exclusive jets with a BDT analysis in each jet is almost equivalent to a BDT where n_{jets} is a variable. The cuts based analysis is comparable to a BDT analysis. The FWM analysis seems to be slightly better in the range $\epsilon_S \approx 0.4 - 0.7$. However, in [33] we find that to proper quantify the improvement in either S/B or $S/\sqrt{S+B}$, we need to exactly quantify the background rejection rates because small changes there can have a strong impact on S/B. For comparison we chose three different signal efficiencies. First $\epsilon_S = 0.232$: here we find the highest significance in our FWM analysis. Second $\epsilon_S = 0.302$: this is the working point for the traditional jet veto analysis. This point is defined by

$$|\Delta y_{j_1 j_2}| > 4.4, \qquad y_{j1} \cdot y_{j2} < 0, \qquad m_{j_1 j_2} > 600 \text{ GeV} . \tag{2.87}$$

And last $\epsilon_S = 0.400$ which is a default point for comparison. Our results are summarized in Tab. 2.10. We find the BDT to be always slightly better than the cuts algorithm. This behavior is expected because a BDT always maps the signal region better. For the jet veto we observe that

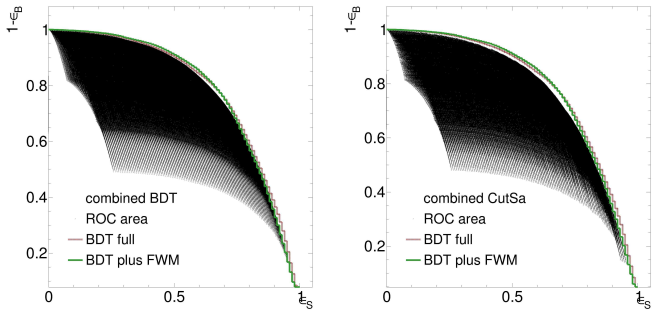

Figure 2.31: Comparison of FWM BDT (solid, green) and n_{jets} BDT (dashed, red). The black area corresponds to the combination of the exclusive jet bin analyses via Eq. (2.86). Left: exclusive jet bin BDT analyses. Right: exclusive jet bin cuts analyses.

ϵ_S	cuts comb.	BDT comb.	BDT n_{jets}	BDT FWM	jet veto	cuts two jet	BDT two jet
	$1-\epsilon_B$ for Δy-selection and S/B						
0.232	0.977(2)	0.981(2)	0.984	0.986	-	0.978	0.981
	0.052(5)	0.063(6)	0.074	0.085	-	0.054	0.063
0.302	0.964(2)	0.965(2)	0.970	0.975	-	0.963	0.967
	0.043(2)	0.044(3)	0.052	0.062	-	0.042	0.047
train eq. (2.87)	-	-	-	-	0.967	0.965	0.967
	-	-	-	-	0.047	0.044	0.047
0.400	0.934(2)	0.944(2)	0.946	0.952	-	0.932	0.942
	0.031(1)	0.037(1)	0.038	0.043	-	0.030	0.035

Table 2.10: Signal efficiencies and background rejection rates as well as S/B for the different analysis strategies discussed in the text.

the cuts in Eq. (2.87) are already optimal. Before we answer our question from the beginning let us address some details of the TMVA training procedure and estimate the uncertainty related to our combination procedure. First of all note that by dividing our analysis in exclusive jet bins the samples to train and test are statistically more limited than in [33][43]. Furthermore, in [33] we are able to include the signal and background cross sections ($\sigma_{\text{WBF}} = 3.78$ pb, $\sigma_{\text{GF}} = 1.71$ pb $\sigma_{\text{WBF}} = 736.2$ pb) as weights into the training process. Due to the limited statistic in the exclusive samples we run into a convergence problem for the cut algorithms and have to set these weights to one. We do not expect this to produce a very big difference in the results. However, knowing which background is more important, we expect the TMVA to find at least better signal areas. This means that especially slight differences (say of $\mathcal{O} < 0.001$) between Tab. 2.10 and [33] might not be significant, but only origin in numerical limitations[44]. Furthermore, our construction of the ROC areas is, of course, not dense. This means that there is an uncertainty connected with the statement what the highest possible background rejection is. We expect this uncertainty to be much more dominant than any uncertainties coming from any possible variations due to statistics. Note, however that this uncertainty is still way smaller than the difference to the other ROC curves, see for example Fig. 2.32. The periodic structure we observe in Fig. 2.32 is a result of the finite subset of points used in Eq. (2.86). We assume that the dips in this structure give a reasonable error estimate for Eq. (2.86)[45]. In terms of background rejection they are of $\mathcal{O}(0.002)$. This leads to an uncertainty for (S/B) for the combined analyses of $\mathcal{O}(0.001)$ for "small" background rejection rates, e.g. the column "cuts combined" in Tab. 2.10 at $\epsilon_S = 0.4$, up to $\mathcal{O}(0.006)$ for "high" background rejection rates, e.g. the column "BDT combined" at $\epsilon_S = 0.232$.

Tab. 2.10 tells us that the boost from the FWM does not result solely from the use of a BDT. We see this especially in the difference between column three and four, for example in row two. There the BDT has access to the different jet bins, but hardly outperforms the classical jet veto

[43] For the two and three jet category we use 100000 signal events for training and 200000 for testing. The backgrounds have 75000 training and 175000 testing events each. However the four jet category only has 35000 training and 65000 testing events for signal as well as backgrounds.

[44] Note also that these numbers are rounded, of course.

[45] From Fig. 2.32 one could draw the conclusion that the deviation in the "up" direction, which is the important one in our analysis, is much smaller than 0.002. However, we like to interpret these uncertainties in a conservative way and take them face valued.

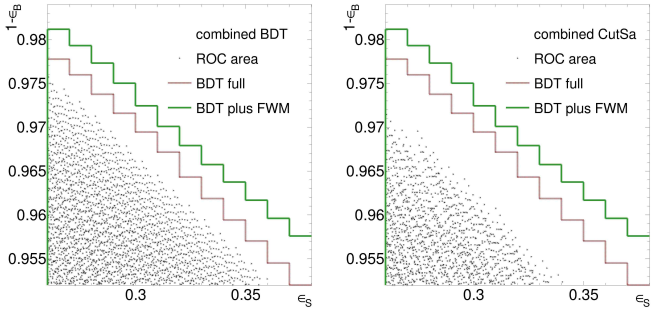

Figure 2.32: Zoom on the ROC curves, respectively area. The black area represents the combination of the BDT (left) and the cut (right) analyses. The dashed, red line is the ROC curve for the full n_{jets} BDT analysis while the solid green curve is from the FWM analysis.

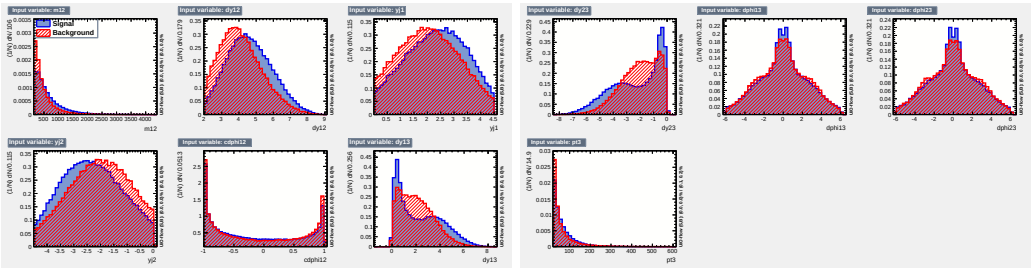

Figure 2.33: Input variables for the exclusive three jet bin. The first five plots in the left figure display Eq. (2.83). The others show the additional observables due to a third jet, namely: $\Delta\eta_{j1j3}$, $\Delta\eta_{j2j3}$, $\Delta\phi_{j1j3}$, $\Delta\phi_{j2j3}$ and $p_{T,j3}$.

with only a 10% gain in S/B. However, using FWMs results in a 32% gain. This means the FWMs include three times more additional information than naively including all jet bins into our analysis. If we move to row one we find that using a full n_{jets} BDT improves the exclusive two jet BDT by 17% while the FWMs gain 35%. This is still a relative factor of two. Where does this additional improvement come from? The FWM built a complete base of observables. Including an additional jet into the analysis does not only change the distribution of the observables of eq. (2.83), but also introduces new observables, e.g. the angels between the tagging jets and the third jet and its p_T. To demonstrate this point we redo the exclusive three jet analysis, but this time include more observables. The input variables are shown in Fig. 2.33. It is immediately clear that $\Delta\eta_{j1j3}$ and $\Delta\eta_{j2j3}$ have strongly discriminating features. Furthermore, we observe a structure where the signal encapsulates the background. This means that we are not able to access this additional information via boxed cuts. We compute ROC curves for this new set of observables and compare them to the old analysis where we only include Eq. (2.83) in Fig. 2.34. As expected the cuts analysis (green, solid respectively dashed) is nearly identical[46]. However, the BDT analysis (thick solid green) shows a huge improvement. This improvement is exactly the source of the improvement in Sec. 2.2. If we also include a fourth jet the set of observables increases accordingly. This is the strength of the FWMs, we do not need to care about the jet spectrum or the dimensionality of the final state too much. In a rather simple final state configuration like WBF we might be able to identify all relevant observables intuitively, but this for sure is not possible anymore for complex final states like $t\bar{t}H$. Furthermore, the dimensonality of the observable space grows rapidly. A case where we cannot rely on simple cut strategies anymore. In this regard FWMs together with a BDT are an elegant solution to efficiently explore high-multiplicity final states.

[46] We also include the "standard" cuts analysis (black) which has even worse convergence problems. This is due to the fact that we randomly probe a 10D space with order 100000 MC points

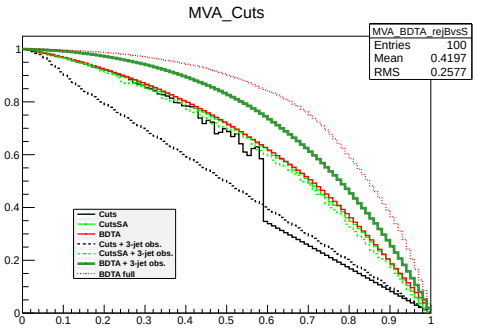

Figure 2.34: ROC curves for the exclusive three-jet analysis. Light green displays the cuts algorithm for eq. (2.83) (solid) plus the new observables (dashed). The BDT analysis is red for eq. (2.83) while the solid thick dark green is a BDT trained on all observables. (Sorry for this ugly plot. Our hard drive died and exactly the data for this plot didn't survive. Thus I had to use this backup version)

Conclusions

In this thesis we have computed the resummed jet spectrum to all orders in the democratic limit. It is the first time that this observable is computed analytically. We find that it obeys the staircase scaling pattern. The consequence is that the ratios of successive jet cross sections are constant. This is a solid first principle QCD statement. The impact of this fact is huge. It allows us to deduct the shape of standard model multi jet backgrounds from the low multiplicity region. We show how this can be used to construct a data driven background estimation. In addition this is the region where higher order calculations are manageable. Once the jet cross sections are fixed we have the full power of multi jet observables to our disposal. We give two examples, which show their usefulness: (1) a fairly model independent dark matter search and (2) the improvement of the WBF Higgs analysis.

The analytic computation is necessary, because exclusive multi jet observables are plagued by huge theory uncertainties. We solve this problem by proving that higher jet multiplicities can be fixed via the staircase scaling argument. However, there is still room for improvement.

- A further study on the choice of the factorization scale within the Sherpa framework could reveal, if scale uncertainties are really a tuning parameter or if there is some (maybe systematic) analysis dependence. Especially in the light of the Fox–Wolfram–Moments this has still to be tested.

- The expansion of the generating functional offers the possibility to study higher order staircase scaling breaking terms. It is left to future research to show their impact. Together with a better analytic understanding of phase space effects scaling features may be useful in substructure studies, too.

At the moment Valerie Lang from the Heidelberg ATLAS group studies scaling features in W plus jets in data. We are very excited to see her results.

Acknowledgments

First of all I want to thank Tilman for giving me this great opportunity. I only start to realize how much support I got in Heidelberg. I hope I can turn the skills and connections obtained here into more interesting research in the future. In this regard I also want to thank Werner Rodejohann representative of the whole IMPRS for *Precision Tests of Fundamental Symmetries* and their never ending support. Let me also thank my collaborators and colleagues here at the ITP: Christoph Englert, Steffen Schumann, Erik Gerwick, David Lopez-Val, Dorival Gonzales-Netto, Daniel Wiegand, Cathrine Bernaciak, Jamie Tattersall, Martin Jankoviak, Felix Kling, and many more. We worked and discussed together and I am grateful for everything I learned so far. The other part in our research program is simulation. I would like to thank Elmar Bittner and Andreas Nussbaumer for local computing support and rescuing our hard drive. I also would like to thank Sabine Richling representative of the bwGRiD, who never got angry at me when going over their quota. My thanks extend to Hans-Christian Schultz-Coulon, Jan Pawlowski, and Björn-Malte Schäfer for reading my thesis and to agree to be my examiners at my PhD defense. An additional thank to Jamie Tattersal and Erick Gerwick for proofreading some parts of this thesis. With all my heart I thank Maria Walch for her never ending support of chocolate, encouragement, and love.

Appendix A
Splitting Kernels

Splitting kernels have first been discussed in the famous paper of Altarelli and Parisi [39], where they compute massless QCD branchings. The procedere developed there can also be used to compute splittings involving massive partons and SUSY vertices. Based on the massless definition given in their paper, we compute the general massive splitting kernels. These can easily be taken to the $m \to 0$ limit. The most important part of the procedere is to notice, that we work in the collinear limit, involved in terms of a p_T expansion. There is one subtelty when working in the collinear limit this way. As for the factorization theorem the equation only holds when we are exactly in the limiting case. What the method actually does is just squaring a vertex. However, a three particle vertex cannot be on shell and conserve energy momentum. Therefor, we have to take special care of the polarization sum. For external gluons we only can allow transverse polarizations. Usually we use the Lorentz structure of the gluon propagator. There the Ward identities garantue that all non-transverse components give zero. However, to use the Ward identities we need energy momentum conservation *and* that all external particles are on shell. The latter condition is given for the four-momenta in the way we expand them in terms of p_T. The former is not. This means we must envorce a transverse polarization sum. We do this by using

$$\sum_{\text{polarizations}} \epsilon^{*\mu}(k)\epsilon^{\nu}(k) = \delta^{mn} - \frac{k^m k^n}{\vec{k}^2}, \tag{A.1}$$

which explicitly removes unphysical modes. Here $(m, n) \in [1, 2, 3]$, while $(\mu, \nu) \in [0, 1, 2, 3]$ and k is the momentum of the gluon. To avoid this issue one can compute the Catanai–Seymour dipoles. These contain the splitting kernels as a limiting case. Some of the techniques shown in the following are also useful for their computation. In terms of physics the dipoles are preferable, because they do conserve energy and momentum by taking a spectator parton into account. The spectator is connected to the splitting kernel via a color line and absorbes the momentum. However, this technique produces more combinatorics and in addition the generating functions are phrased in terms of standard splitting kernels. Therefor, we limit ourselves to the standard splitting kernels here. Note, however, that modern parton showers do use dipoles instead of splitting kernels. It is left to future work to study if dipoles are useful for generating functions, too.

To compute the actual kernels of the type $i \to jk$ we phrase the four momenta of the involved particles in terms of energy fraction $z = E_j/E_i$ and transverse momentum with respect to the direction of i. These momenta are on shell, but conserve energy and momentum only up to terms of $\mathcal{O}(p_T^2)$.

$$k_i = \begin{pmatrix} E \\ P \\ 0 \end{pmatrix}, \quad k_j = \begin{pmatrix} zE + \frac{p_T^2 - z^2 m_i^2 + m_j^2}{4zE} \\ zP - \frac{p_T^2 - z^2 m_i^2 + m_j^2}{4zP} \\ p_T \end{pmatrix}, \quad k_k = \begin{pmatrix} (1-z)E + \frac{p_T^2 - (1-z)^2 m_i^2 + m_k^2}{4(1-z)E} \\ (1-z)P - \frac{p_T^2 - (1-z)^2 m_i^2 + m_k^2}{4(1-z)P} \\ -p_T \end{pmatrix}. \tag{A.2}$$

A technical note: Invovling masses we speak of the quasi collinear limit contrasted to the collinear limit for massles partons. The collinear divergence will be shielded by these mass terms in form of terms of $\log p_T/m$. The limiting case is defined as haveing p_T and all m approach zero the same way. Only the leading terms in this expansion are kept in Eq. (A.2). A consequence of this is that the following relation holds.

$$(k_j + k_k)^2 = \frac{p_T^2}{z(1-z)} + \frac{m_j^2}{z} + \frac{m_k^2}{1-z}. \tag{A.3}$$

Writing the momenta in this specific way the computation of the splitting kernels reduces to

$$P_{i\to jk} = \frac{1}{2} \frac{z(1-z)}{(k_j + k_k)^2 - m_i^2} \sum_{\text{polarizations}} |V_{i\to jk}|^2, \tag{A.4}$$

where V is the vertex of the branching as defined by the Feynman rules. We compute LO vertices here. It is possible to go beyond this approximation. Higher order corrections result in $1 \to 3$ splitting configurations and will in the end result in better logarithmic precission.

There are many different ways to obtain the splitting kernels. We follow the conventions by Catani et al. The original method uses slightly different four momenta, but yields the same results. Another method, however, widely used in text books [22,24,25,166] uses a collinear approximation for the external spinors rather then for the resulting product of four momenta. While this works in the massless case it is useless for the massive. In the end of this appendix we compare the different ways one can use spinor techniques and what the trouble for the massive case is. In addition the computation reviewed here are the first step to also understand the Catani Seymour dipoles, which are very important quantities in resummation as well as next to leading order computations.

A.1 Massless splitting kernels

In the limit $m \to 0$ Eq. (A.4) reduces to

$$P_{i\to jk} = \frac{1}{2}\frac{z(1-z)}{p_T^2} \sum_{\text{polarizations}} |V_{i\to jk}|^2 . \tag{A.5}$$

The four momenta are now defined as

$$k_i = \begin{pmatrix} E \\ E \\ 0 \end{pmatrix}, \quad k_j = \begin{pmatrix} zE + \frac{p_T^2}{4zE} \\ zE - \frac{p_T^2}{4zE} \\ p_T \end{pmatrix}, \quad k_k = \begin{pmatrix} (1-z)E + \frac{p_T^2}{4(1-z)E} \\ (1-z)E - \frac{p_T^2}{4(1-z)E} \\ -p_T \end{pmatrix}. \tag{A.6}$$

This yields the following list of products of four and three momenta

$$\begin{aligned}
k_i k_k &= \frac{p_T^2}{2(1-z)}, \\
k_j k_k &= \frac{p_T^2}{2z(1-z)}, \\
\vec{k}_i \vec{k}_j &= zE^2 - \frac{p_T^2}{4z}, \\
\vec{k}_i \vec{k}_k &= (1-z)E^2 - \frac{p_T^2}{4(1-z)}, \\
\vec{k}_j \vec{k}_k &= z(1-z)E^2 - \frac{p_T^2}{4z(1-z)}(1 + 2z(1-z)) + \mathcal{O}(p_T^4), \\
\vec{k}_i^2 &= E^2, \\
\vec{k}_j^2 &= \frac{1}{2}p_T^2 + z^2E^2 + \mathcal{O}(p_T^4), \\
\vec{k}_k^2 &= \frac{1}{2}p_T^2 + (1-z)^2E^2 + \mathcal{O}(p_T^4).
\end{aligned} \tag{A.7}$$

These are the products which appear in the calculation of the splitting kernels. Thus all we have to do is write down the vertex, square it, and put these resluts in.

Gluon radiation from a quark

The matrix element is

$$i\mathcal{M} = ig_s \bar{u}(k_k) T^j \gamma_\mu u(k_i) \epsilon^{*\mu}. \tag{A.8}$$

Squaring and averaging over spins and colors yields

$$\begin{aligned}\sum_{\text{polarizations}} |V_{i\to jk}|^2 &= \frac{1}{2} C_F g_s^2 tr\{\not{k}_k \gamma_\mu \not{k}_i \gamma_\nu\} \epsilon^{*\mu} \epsilon^{\nu} \\ &= C_F g^2 2 \left(k_{k,\mu} k_{i,\nu} - k_k k_i g_{\mu\nu} + k_{i,\mu} k_{k,\nu}\right) \left(\delta^{mn} - \frac{k_j^m k_j^n}{\vec{k}_j^2}\right) \\ &= 2 C_F g^2 \left(\left(2\vec{k}_i \vec{k}_k + 3 k_i k_k\right) - \left(2 \frac{(\vec{k}_i \vec{k}_j)(\vec{k}_j \vec{k}_k)}{\vec{k}_B^2} + k_i k_k\right)\right). \end{aligned} \tag{A.9}$$

We therefor have in leading order of p_T^2

$$\begin{aligned} P_{q\to qg} &= C_F \frac{1+(1-z)^2}{z}, \\ P_{q\to gq} &= C_F \frac{1+z^2}{1-z} \quad \text{(by kinematic symmetry)}, \\ P_{\tilde{g}\to \tilde{g}g} &= C_A \frac{1+(1-z)^2}{z} \quad \text{(by super symmetry)}, \\ P_{\tilde{g}\to g\tilde{g}} &= C_A \frac{1+z^2}{1-z}. \end{aligned} \tag{A.10}$$

Gluon to quark splitting

The matrix element is

$$i\mathcal{M} = ig\bar{u}(k_j) T^i \gamma_\mu v(k_k) \epsilon^\mu(k_i). \tag{A.11}$$

Squaring and averaging over polarizations and colors yields

$$\begin{aligned}\sum_{\text{polarizations}} |V_{i\to jk}|^2 &= 2 T_R \left(k_{j,\mu} k_{k,\nu} - k_j k_k g_{\mu\nu} + k_{k,\mu} k_{j,\nu}\right) \left(\delta^{mn} - \frac{k_i^m k_i^n}{\vec{k}_i^2}\right) \\ &= 2 T_R g^2 \left(\left(2\vec{k}_j \vec{k}_k + 3 k_j k_k\right) - \left(2 \frac{(\vec{k}_i \vec{k}_j)(\vec{k}_i \vec{k}_k)}{\vec{k}_i^2} + k_j k_k\right)\right). \end{aligned} \tag{A.12}$$

We therefor have in leading order of p_T^2

$$\begin{aligned} P_{g\to q\bar{q}} &= T_R \left(z^2 + (1-z)^2\right), \\ P_{g\to \tilde{g}\tilde{g}} &= C_A \left(z^2 + (1-z)^2\right) \quad \text{(by super symmetry)}. \end{aligned} \tag{A.13}$$

Gluon triple vertex

The matrix element is

$$i\mathcal{M} = -g f^{ijk} \left(\left(k_i - k_j\right)_\rho g_{\mu\nu} + \left(k_j - k_k\right)_\mu g_{\nu\rho} + \left(k_k - k_i\right)_\nu g_{\rho\mu}\right) \epsilon^\mu(k_i) \epsilon^{*\nu}(k_j) \epsilon^{*\rho}(k_k). \tag{A.14}$$

Squaring and averaging over polarizations and colors yields

$$\sum_{\text{polarizations}} |V_{i\to jk}|^2 = \frac{8 p_T^2 (1 - 2z + 3z^2 - 2z^3 + z^4)}{(1-z)^2 z^2} + \mathcal{O}(p_T^4). \tag{A.15}$$

We therefor have in leading order of p_T^2

$$P_{g\to gg} = C_A \left(\frac{1-z}{z} + \frac{z}{1-z} + z(1-z)\right). \tag{A.16}$$

Squark gluino gluon vertex

We start with the squark in the initial state. The matrix element is

$$i\mathcal{M} = -\mathbb{1}_L \bar{u}(k_j) i g \sqrt{2} T^j \mathbb{P}_R v(-k_k). \tag{A.17}$$

Squaring and averaging over colors yields

$$\begin{aligned}
\mathcal{M}^2 &= C_F g^2 2tr\{\not{k}_j \mathbb{P}_R \not{k}_k \mathbb{P}_L\} \\
&= 2C_F g^2 \frac{1}{4} \left(4k_j k_k - tr\{\not{k}_j \gamma^5 \not{k}_k \gamma^5\}\right) \\
&= 2C_F g^2 \frac{1}{4} \left(4k_j k_k + 4k_j k_k\right) \\
&= 4C_F g^2 k_j k_k.
\end{aligned} \tag{A.18}$$

We therefor have in leading order of p_T^2

$$P_{\tilde{q}\to\tilde{g}q}(z) = C_F \times (1). \tag{A.19}$$

Putting the quark in the initial state we have $k_k \to k_i$. This yields

$$P_{q\to\tilde{q}\tilde{g}}(z) = C_F(1-z). \tag{A.20}$$

Symmetry then yields

$$P_{q\to\tilde{g}\tilde{q}}(z) = C_F(z). \tag{A.21}$$

With the gluino in the initial state only the color factor changes $C_F \to T_R$.

$$\begin{aligned}
P_{\tilde{g}\to\tilde{q}q}(z) &= T_R(1-z). \\
P_{\tilde{g}\to q\tilde{q}}(z) &= T_R(z).
\end{aligned} \tag{A.22}$$

Observe that, because no gauge coupling is involved, no configuration produces a soft divergence.

A.2 Massive splitting kernels

The computation for the massive splitting kernels works analoge. Only two things change. For spin sums we need to put in the massive Dirac equation and squared four momenta might yield a mass term now. The list of import four and three momentum products depends therefor on the particular configuration we study.

Gluon radiation from a heavy quark

For this kernel we have $m_i = m_j = m_Q$ and $m_k = 0$. The matrix element is changed and includes now a factor m_Q^2. The vector products are:

$$\begin{aligned}
k_i k_j &= \frac{p_T^2 + m_Q^2(1+z^2)}{2z} \\
\vec{k}_i \vec{k}_j &= zP^2 - \frac{p_T^2 + (1-z^2)m_Q^2}{4z} \\
\vec{k}_i \vec{k}_k &= (1-z)P^2 - \frac{p_T^2 - (1-z)^2 m_Q^2}{4(1-z)} \\
\vec{k}_j \vec{k}_k &= z(1-z)P^2 - \frac{(z^2+(1-z)^2)p_T^2 + (1-z)^2(1-2z)m_Q^2}{4z(1-z)} \\
\vec{k}_k^2 &= \frac{p_T^2}{2} + (1-z)^2 \left(P^2 + \frac{m_Q^2}{2}\right).
\end{aligned} \tag{A.23}$$

For the squared matrix element we find

$$\begin{aligned}
&4g^2\left(\vec{k}_i\vec{k}_j + k_ik_j - m_Q^2 - \frac{(\vec{k}_i\vec{k}_k)(\vec{k}_j\vec{k}_k)}{\vec{k}_k^2}\right)\\
&= 2g^2\frac{p_T^2+(1-z)^2m_Q^2}{z(1-z)}\left(\frac{1+z^2}{1-z} - \frac{2z(1-z)m_Q^2}{p_T^2+(1-z)^2m_Q^2}\right).
\end{aligned} \tag{A.24}$$

We therefor have in leading order of p_T^2

$$\begin{aligned}
P_{Q\to Qg} &= C_F\left(\frac{1+z^2}{1-z} - \frac{2m_Q^2}{(p_Q+p_g)^2}\right),\\
P_{Q\to Qg} &= C_F\left(\frac{1+(1-z)^2}{z} - \frac{2m_Q^2}{(p_g+p_Q)^2}\right) \qquad \text{(by symmetry)}.
\end{aligned} \tag{A.25}$$

Gluon splitting to two heavy quarks

For this kernel we have $E = P \to m_i = 0, m_j = m_k = m_Q$. The matrix element is changed and includes now a factor $2m_Q^2$. The vector products are:

$$\begin{aligned}
k_jk_k &= \frac{p_T^2+m_Q^2}{2z(1-z)} - m_Q^2\\
\vec{k}_j\vec{k}_k &= m_Q^2 + z(1-z)P^2 - \frac{p_T^2+m_Q^2}{4z(1-z)}\left(1+2z(1-z) - \frac{p_T^2+m_Q^2}{4P^2}\right)\\
\vec{k}_i\vec{k}_j &= zP^2 - \frac{p_T^2+m_Q^2}{4z}\\
\vec{k}_i\vec{k}_k &= (1-z)P^2 - \frac{p_T^2+m_Q^2}{4(1-z)}\\
\vec{k}_i^2 &= P^2
\end{aligned} \tag{A.26}$$

For the squared matrix element we find

$$\begin{aligned}
&2T_Rg^2\left(\left(2\vec{k}_j\vec{k}_k + 3k_jk_k + 3m_Q^2\right) - \left(2\frac{(\vec{k}_i\vec{k}_j)(\vec{k}_i\vec{k}_j)}{\vec{k}_i^2} + k_jk_k + m_Q^2\right)\right)\\
&= 4T_Rg^2\left(\frac{p_T^2+m_Q^2}{2z(1-z)} + m_Q^2 + z(1-z)P^2 - \frac{p_T^2+m_Q^2}{4z(1-z)}\left(1+2z(1-z) - \frac{p_T^2+m_Q^2}{4P^2}\right)\right.\\
&\left. - \frac{\left(zP^2 - \frac{p_T^2+m_Q^2}{4z}\right)\left((1-z)^2P^2 - \frac{p_T^2+m_Q^2}{4(1-z)}\right)}{P^2}\right)\\
&= 4g^2T_R\frac{p_T^2+m_Q^2}{2z(1-z)}\left[1-2z(1-z) + \frac{2z(1-z)m_Q^2}{p_T^2+m_Q^2}\right].
\end{aligned} \tag{A.27}$$

We therefor have in leading order of p_T^2

$$P_{g\to Q\bar{Q}} = T_R\left(1-2z(1-z) + \frac{2m_Q^2}{(p_Q+p_{\bar{Q}})^2}\right) \tag{A.28}$$

Massive squark gluino gluon vertex

With the squark in the initial state the squared matrix element does not change because the projectors in the vertex multiply to zero for the mass term. However, the product of four momenta does change, because of our new parametrization. We now have

$$2k_jk_k = (k_j+k_k)^2 + m_i^2 - m_j^2 - m_k^2 - m_i^2. \tag{A.29}$$

Therefor, we get

$$P_{\tilde{q}\to\tilde{g}Q} = C_F\left(1-\frac{m_{\tilde{g}}^2+m_Q^2-m_{\tilde{q}}^2}{(k_j+k_k)^2-m_{\tilde{q}}^2}\right),$$
$$P_{qQ\to\tilde{g}\tilde{q}} = C_F\left(1-z-\frac{m_{\tilde{g}}^2+m_{\tilde{q}}^2-m_Q^2}{(k_j+k_k)^2-m_Q^2}\right),$$
$$P_{\tilde{g}\to\tilde{q}Q} = T_R\left(1-z-\frac{m_{\tilde{q}}^2+m_Q^2-m_{\tilde{g}}^2}{(k_j+k_k)^2-m_{\tilde{g}}^2}\right). \tag{A.30}$$

A.3 Spinors and the collinear limit

There is another way to obtain the splitting kernels using a collinear representation of the external spinors and polarization vectors. In the literature this is done for massless splitting kernels. However, many different parametrizations are floating around. To understand the connection between the different methods we study general massive splitting kernels in different Dirac representations and their transformation properties under a basis change. We then compute massless spinors in the collinear limit. We explicitly check the connection between the different basis choices and how they are connected to the collinear spinors used in the literature.

Spinors in Weyl basis: Srednicki

In chapter 38 spinors are introduced by solving the Dirac equation in the Weyl basis.

$$(\not{p}+m)u_{W,s}(\mathbf{p}) = 0. \tag{A.31}$$

The Weyl basis is defined by the following representation of gamma matrices

$$\gamma_W^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix} \quad \gamma_W^5 = \begin{pmatrix} -\mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix}. \tag{A.32}$$

In the rest frame where $\mathbf{p}=0$ we get the following solutions

$$u_{W,+}(0) = \sqrt{m}\begin{pmatrix}1\\0\\1\\0\end{pmatrix} \quad u_{W,-}(0) = \sqrt{m}\begin{pmatrix}0\\1\\0\\1\end{pmatrix} \quad v_{W,+}(0) = \sqrt{m}\begin{pmatrix}0\\1\\0\\-1\end{pmatrix} \quad v_{W,-}(0) = \sqrt{m}\begin{pmatrix}-1\\0\\1\\0\end{pmatrix}. \tag{A.33}$$

To boost to an arbitrary frame $\mathbf{p}$ we use

$$D_W(\Lambda) = \exp\left(i\eta\hat{\mathbf{p}}\mathbf{K_W}\right) = \cosh\left(\frac{\eta}{2}\right)\ \mathbb{1} + \sinh\left(\frac{\eta}{2}\right)\begin{pmatrix}-\sigma\hat{\mathbf{p}} & 0\\ 0 & \sigma\hat{\mathbf{p}}\end{pmatrix}$$
$$= \begin{pmatrix} \cosh\left(\frac{\eta}{2}\right)-\cos\theta\,\sinh\left(\frac{\eta}{2}\right) & -\sin\theta\,\sinh & 0 & 0 \\ -\sin\theta\,\sinh\left(\frac{\eta}{2}\right) & \cosh\left(\frac{\eta}{2}\right)+\cos\theta\,\sinh\left(\frac{\eta}{2}\right) & 0 & 0 \\ 0 & 0 & \cosh\left(\frac{\eta}{2}\right)+\cos\theta\,\sinh\left(\frac{\eta}{2}\right) & \sin\theta\,\sinh\left(\frac{\eta}{2}\right) \\ 0 & 0 & \sin\theta\,\sinh\left(\frac{\eta}{2}\right) & \cosh\left(\frac{\eta}{2}\right)-\cos\theta\,\sinh\left(\frac{\eta}{2}\right) \end{pmatrix}. \tag{A.34}$$

where $K_W^j = i/4[\gamma_W^j,\gamma_W^0]$ and η is the rapidity. Note that we use $\phi=0$. We thus have $\cosh\left(\frac{\eta}{2}\right) = \sqrt{E+m}/\sqrt{2m} = A/\sqrt{2m}$ and $\sinh\left(\frac{\eta}{2}\right) = \sqrt{E-m}/\sqrt{2m} = B/\sqrt{2m}$. The boost matrix now reads

$$\frac{1}{\sqrt{2m}}\begin{pmatrix} A-B\cos\theta & -B\sin\theta & 0 & 0 \\ -B\sin\theta & A+B\cos\theta & 0 & 0 \\ 0 & 0 & A+B\cos\theta & B\sin\theta \\ 0 & 0 & B\sin\theta & A-B\cos\theta \end{pmatrix}. \tag{A.35}$$

Using the rest frame solution from above general massive Dirac spinors in Weyl represantation read

$$u_{W,+}(\mathbf{p}) = \frac{1}{\sqrt{2}}\begin{pmatrix} A - B\cos\theta \\ -B\sin\theta \\ A + B\cos\theta \\ B\sin\theta \end{pmatrix} \quad u_{W,-}(\mathbf{p}) = \frac{1}{\sqrt{2}}\begin{pmatrix} -B\sin\theta \\ A + B\cos\theta \\ B\sin\theta \\ A - B\cos\theta \end{pmatrix}$$
$$v_{W,+}(\mathbf{p}) = \frac{1}{\sqrt{2}}\begin{pmatrix} -B\sin\theta \\ A + B\cos\theta \\ -B\sin\theta \\ -A + B\cos\theta \end{pmatrix} \quad v_{W,-}(\mathbf{p}) = \frac{1}{\sqrt{2}}\begin{pmatrix} -A + B\cos\theta \\ B\sin\theta \\ A + B\cos\theta \\ B\sin\theta \end{pmatrix}. \tag{A.36}$$

Spinors in Dirac basis: Bjorken and Drell

In chapter 3 spinors are introduced by solving the Dirac equation in Dirac representation.

$$\gamma_D^0 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} \quad \gamma_D^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad \gamma_D^5 = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{A.37}$$

The solution in the rest frame is

$$u_{D,+}(0) = \sqrt{2m}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad u_{D,-}(0) = \sqrt{2m}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad v_{D,+}(0) = \sqrt{2m}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad v_{D,-}(0) = \sqrt{2m}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{A.38}$$

In contrast to Bjorken and Drell we introduce the norm $\sqrt{2m}$ to avoid mass singularities in general spinors. The transformation to an arbitry frame is given by

$$\begin{aligned} D_D(\Lambda) = S &= \exp\left(-\frac{\eta}{2}\frac{\sigma\mathbf{v}}{|\mathbf{v}|}\right) \\ &= \sqrt{\frac{E+m}{2m}}\begin{pmatrix} 1 & 0 & \frac{p_z}{E+m} & \frac{p_x}{E+m} \\ 0 & 1 & \frac{p_x}{E+m} & \frac{-p_z}{E+m} \\ \frac{p_z}{E+m} & \frac{p_x}{E+m} & 1 & 0 \\ \frac{p_x}{E+m} & \frac{-p_z}{E+m} & 0 & 1 \end{pmatrix}. \end{aligned} \tag{A.39}$$

Using the identy $p_{z/x} = \sqrt{E^2 - m^2}\cos/\sin\theta$ and introducing the terms A and B as above we get

$$= \sqrt{\frac{1}{2m}}\begin{pmatrix} A & 0 & B\cos\theta & B\sin\theta \\ 0 & A & B\sin\theta & -B\cos\theta \\ B\cos\theta & B\sin\theta & A & 0 \\ B\sin\theta & -B\cos\theta & 0 & A \end{pmatrix}. \tag{A.40}$$

For simplicity we set $p_y = 0$ corresponding to $\phi = 0$ above. Because of the structure of the rest frame spinors we immediately read of the general form in Dirac representation.

$$u_{D,+}(\mathbf{p}) = \begin{pmatrix} A \\ 0 \\ B\cos\theta \\ B\sin\theta \end{pmatrix} \quad u_{D,-}(\mathbf{p}) = \begin{pmatrix} 0 \\ A \\ B\sin\theta \\ -B\cos\theta \end{pmatrix}$$
$$v_{D,+}(\mathbf{p}) = \begin{pmatrix} B\cos\theta \\ B\sin\theta \\ A \\ 0 \end{pmatrix} \quad v_{D,-}(\mathbf{p}) = \begin{pmatrix} B\sin\theta \\ -B\cos\theta \\ 0 \\ A \end{pmatrix}. \tag{A.41}$$

Basis transformation: Pokorski

In appendix A we find the unitary transformation connecting the Weyl and the Dirac basis. The transformation matrix reads

$$\begin{aligned}
\Psi_{\mathrm{D}} &= U^\dagger \Psi_{\mathrm{W}} \\
U &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \\
U^\dagger &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \\
U^\dagger U &= \mathbb{1}.
\end{aligned} \tag{A.42}$$

For the Weyl spinors from Srednicki we get

$$\begin{aligned}
u_{D,+} = U^\dagger u_{W,+} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} A - B\cos\theta \\ -B\sin\theta \\ A + B\cos\theta \\ B\sin\theta \end{pmatrix} \\
&= \begin{pmatrix} A \\ 0 \\ B\cos\theta \\ B\sin\theta \end{pmatrix}.
\end{aligned} \tag{A.43}$$

which is the same expression as in Bjorken and Drell. For the boost matrix we have

$$\begin{aligned}
D_D &= U^\dagger D_W U \\
&= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2m}} \begin{pmatrix} A - B\cos\theta & -B\sin\theta & 0 & 0 \\ -B\sin\theta & A + B\cos\theta & 0 & 0 \\ 0 & 0 & A + B\cos\theta & B\sin\theta \\ 0 & 0 & B\sin\theta & A - B\cos\theta \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \\
&= \frac{1}{\sqrt{2m}} \begin{pmatrix} A & 0 & B\cos\theta & B\sin\theta \\ 0 & A & B\sin\theta & -B\cos\theta \\ B\cos\theta & B\sin\theta & A & 0 \\ B\sin\theta & -B\cos\theta & 0 & A \end{pmatrix}.
\end{aligned} \tag{A.44}$$

Again we rederive the result from Bjorken and Drell.

Massless limit

In the massless limit $m \to 0$ and thus $A = B = \sqrt{E}$. We are mainly interested in the small scattering angle approxiamtion. Therefor we use $\cos\theta \approx 1$ and $\sin\theta \approx \theta$. In the massless limit the spinors only depend on the particle energy and the scattering angel. In the Dirac basis we find

$$\begin{aligned}
u_{D,+}(E,\theta) &= \sqrt{E} \begin{pmatrix} 1 \\ 0 \\ 1 \\ \theta \end{pmatrix} & u_{D,-}(E,\theta) &= \sqrt{E} \begin{pmatrix} 0 \\ 1 \\ \theta \\ -1 \end{pmatrix} \\
v_{D,+}(E,\theta) &= \sqrt{E} \begin{pmatrix} 1 \\ \theta \\ 1 \\ 0 \end{pmatrix} & v_{D,-}(E,\theta) &= \sqrt{E} \begin{pmatrix} \theta \\ -1 \\ 0 \\ 1 \end{pmatrix}.
\end{aligned} \tag{A.45}$$

The same spinors in Weyl basis read

$$u_{W,+}(E,\theta) = \sqrt{2E}\begin{pmatrix} 0 \\ -\theta/2 \\ 1 \\ \theta/2 \end{pmatrix} \qquad u_{W,-}(E,\theta) = \sqrt{2E}\begin{pmatrix} -\theta/2 \\ 1 \\ \theta/2 \\ 0 \end{pmatrix}$$
$$v_{W,+}(E,\theta) = \sqrt{2E}\begin{pmatrix} -\theta/2 \\ 1 \\ -\theta/2 \\ 0 \end{pmatrix} \qquad v_{W,-}(E,\theta) = \sqrt{2E}\begin{pmatrix} 0 \\ \theta/2 \\ 1 \\ \theta/2 \end{pmatrix}. \tag{A.46}$$

Helicity spinors in Dirac basis: Ellis, Sterling and Webber

In chapter 5 we find for the helicity spinors in Dirac basis in the small angel approximation

$$u_{D,+}(E,\theta) = \sqrt{E}\begin{pmatrix} 1 \\ \theta/2 \\ 1 \\ \theta/2 \end{pmatrix} \qquad u_{D,-}(E,\theta) = \sqrt{E}\begin{pmatrix} -\theta/2 \\ 1 \\ \theta/2 \\ -1 \end{pmatrix}$$
$$v_{D,+}(E,\theta) = \sqrt{E}\begin{pmatrix} -\theta/2 \\ -1 \\ \theta/2 \\ 1 \end{pmatrix} \qquad v_{D,-}(E,\theta) = \sqrt{E}\begin{pmatrix} -1 \\ \theta/2 \\ -1 \\ \theta/2 \end{pmatrix}. \tag{A.47}$$

These spinors do not coincide with those we found in the massless limit. They are actually helicity spinors describing Weyl fermions. To come to Weyl's solution we need to incorporate projection operators

$$\mathbb{P}_{R/L} = \frac{1}{2}\left(\mathbb{1} \pm \gamma^5\right). \tag{A.48}$$

A general Dirac spinor has four independent complex entries, while a Weyl spinor only has two. Acting with $\mathbb{P}$ on the spinors we find two solutions:

$$u_{D,+}(E,\theta) = \sqrt{E}\begin{pmatrix} 1 \\ \theta/2 \\ 1 \\ \theta/2 \end{pmatrix} \qquad u_{D,-}(E,\theta) = \sqrt{E}\begin{pmatrix} -\theta/2 \\ 1 \\ \theta/2 \\ -1 \end{pmatrix}. \tag{A.49}$$

The other solutions we find by constructing explicitly the anti-particles with

$$v_{D,s}(p) = \mathcal{C}\mathcal{P}\bar{u}^T_{D,s}(p), \tag{A.50}$$

where $\mathcal{C} = i\gamma^2\gamma^0$ is the charge conjugation and $\mathcal{P} = i\gamma^0$ & $\theta \to -\theta$ is the parity transformation. We transform an outgoing particle $\bar{u}$ into an incoming anti-particle v. We get

$$v_{D,+}(E,\theta) = i\sqrt{E}\begin{pmatrix} -\theta/2 \\ -1 \\ \theta/2 \\ 1 \end{pmatrix} \qquad v_{D,-}(E,\theta) = i\sqrt{E}\begin{pmatrix} -1 \\ \theta/2 \\ -1 \\ \theta/2 \end{pmatrix}. \tag{A.51}$$

To find the spinors in Weyl basis we use the unitary transformation from above. We find

$$u_{W,+}(E,\theta) = \sqrt{2E}\begin{pmatrix} 0 \\ 0 \\ 1 \\ \theta/2 \end{pmatrix} \qquad u_{W,-}(E,\theta) = \sqrt{2E}\begin{pmatrix} -\theta/2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$
$$v_{W,+}(E,\theta) = \sqrt{2E}\begin{pmatrix} -\theta/2 \\ -1 \\ 0 \\ 0 \end{pmatrix} \qquad v_{W,-}(E,\theta) = \sqrt{2E}\begin{pmatrix} 0 \\ 0 \\ -1 \\ \theta/2 \end{pmatrix}. \tag{A.52}$$

We find a reference to this representation in Peskin and Schroeder chapter 17 where the left handed spinor in the splitting $p \to qk$ is defined as

$$u_L(k) = \sqrt{2k^0} \begin{pmatrix} \frac{p_T}{2(1-z)p} \\ 1 \\ 0 \\ 0 \end{pmatrix}. \tag{A.53}$$

Keeping in mind that $|k| \approx (1-z)p$ and that k is defined with the *minus* p_T component we have

$$p_T \approx (1-z)p \times -\theta. \tag{A.54}$$

Appendix B
Hypothesis test

This appendix has already been published in [27] and was originally worked out by Christoph Englert.

In this section we briefly review the binned log-likelihood ratio hypothesis tests which we apply in Section 1.2. It discriminates between two specific hypotheses and has been used for the combined LEP-Higgs limits [85], Tevatron analyses [86], and in various contexts of LHC Higgs phenomenology [87]. According to the Neyman–Pearson lemma [88] the likelihood ratio is the most powerful test statistic (e.g. signal+background vs background-only). We compute the (binned) log-likelihood ratio

$$\mathcal{Q} = -2\log Q = -2\log\frac{L(\text{data}\,|\,\mathcal{S}+\mathcal{B})}{L(\text{data}\,|\,\mathcal{B})} = 2\left[s - n\log\left(1+\frac{s}{b}\right)\right] \overset{\text{binned}}{=} 2\sum_{i\in\text{bins}}\left[s_i - n_i\log\left(1+\frac{s_i}{b_i}\right)\right], \tag{B.1}$$

where s and b denote the signal $\mathcal{S}$ and background $\mathcal{B}$ event numbers for a given luminosity, split into the bins i. The probabilities to observe n events given the expected numbers s and b in Eq.(B.1) are determined by Poisson distributions

$$L(\text{data}\,|\,\mathcal{S}+\mathcal{B}) = \frac{(b+s)^n e^{-(s+b)}}{n!} \qquad L(\text{data}\,|\,\mathcal{B}) = \frac{b^n e^{-b}}{n!}\,. \tag{B.2}$$

The sum in Eq.(B.1) extends over all contributing channels. The likelihood distributions we generate as pseudo-data around each hypothesis' central value, which means that in principle we can include any kind of correlation. In this work we limit ourselves to statistically independent bins i of the n_{jets} and m_{eff} distributions. The set of entries in each bin $\{n_i\}$ we simulate numerically and histogram them as a function of $\mathcal{Q}$, following the Neyman-Pearson lemma. To simulate the log-likelihood distributions we need to specify which hypothesis the bin entries $\{n_i\}$ should follow, i.e. we can compute $\mathcal{Q}_{\mathcal{S}+\mathcal{B}}$ or $\mathcal{Q}_{\mathcal{B}}$. In Figure B.1 we show both $\mathcal{Q}$ distributions for the binned one-dimensional n_{jets} distribution studied in our paper.

In our analysis we are interested in the probability that the background alone fakes the expected signal+background distributions. This confidence level is given by the integral of the background distribution $\mathcal{Q}_{\mathcal{B}}$ over the signal+background range, indicated by the red-shaded region in Figure B.1. This signal+background range is defined as all likelihood values above the median of the

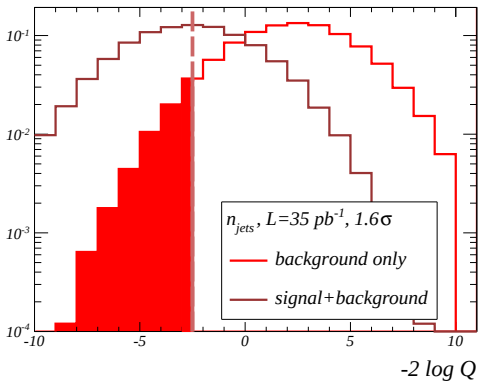

Figure B.1: Log-likelihood ratio distributions based on the n_{jets} discriminator for a luminosity of 35 pb^{-1}. The confidence level is computed by evaluating the overlap of the background-only distribution with the signal+background maximum.

likelihood distribution assuming the signal+background hypothesis

$$\mathrm{CL}_{\mathcal{B}} = \int_{-\infty}^{\langle \mathcal{Q}_{\mathcal{S}+\mathcal{B}} \rangle} \mathrm{d}\mathcal{Q}\, \mathcal{Q}_{\mathcal{B}} = \mathrm{erfc}\left(\frac{\mathcal{Z}}{\sqrt{\in}} \right) , \tag{B.3}$$

where for illustration purposes we convert the confidence levels into the Gaussian number of standard deviations Z via the inverse error function.

Appendix C
Generating Functionals

C.1 The MLLA computation

We demonstrate the computation of the exponentiated form of the generating function for the case of $q \to qg$ splitting at NLL. This means we include the running of the coupling. Eq. (1.11) then reads with the replacement $\mu_F \to t$ and $\mu_0 \to t_0$

$$\Phi_q(t) = u\Delta_q(t) + \int_{t_0}^{t} \frac{dt'}{t'} \frac{\Delta_q(t)}{\Delta_q(t')} \int_0^1 dz \frac{\alpha_s(z^2(1-z)^2t')}{2\pi} C_F \left(\frac{2}{1-z} - \big(1+z\big) \right) \Phi_q(z^2t')\Phi_g((1-z)^2t'). \tag{C.1}$$

In a first step we simplify the divergent part by performing the transformation $(1-z)^2t' = t''$, while setting $z = 1$, where possible (logarithmic expansion). Note that $t'' \in [t', t_0]$ introduces another minus sign.

$$-2\frac{dz}{1-z} = \frac{dt''}{t''}$$

$$\Rightarrow \int_0^1 dz \frac{\alpha_s(z^2(1-z)^2t')}{2\pi} C_F \left(\frac{2}{1-z} \right) \Phi_q(z^2t')\Phi_g((1-z)^2t') = \Phi_q(t') \int_{t_0}^{t'} dt'' \frac{C_F}{2\pi} \frac{\alpha_s(t'')}{t''} \Phi_g(t''). \tag{C.2}$$

For the non-divergent part we use an approximation [25], where the z dependence of the functions is neglected, while all explicit occurrences of z are substituted by its average of $1/2$.

$$\int_0^1 dz \frac{\alpha_s(z^2(1-z)^2t')}{2\pi} C_F \left(-\big(1+z\big) \right) \Phi_q(z^2t')\Phi_g((1-z)^2t') = C_F \frac{\alpha_s(t')}{2\pi} \left(-\frac{3}{2} \right) \Phi_q(t')\Phi_g(t'). \tag{C.3}$$

Hence eq. (C.1) now reads

$$\Phi_q(t) = u\Delta_q(t) + \int_{t_0}^{t} \frac{dt'}{t'} \frac{\Delta_q(t)}{\Delta_q(t')} \Phi_q(t') \frac{C_F}{2\pi} \left[\int_{t_0}^{t'} dt'' \frac{\alpha_s(t'')}{t''} \Phi_g(t'') - \frac{3}{2}\alpha_s(t')\Phi_g(t') \right], \tag{C.4}$$

while its differentiation yields

$$\begin{aligned} \Phi_q'(t) &= u\Delta_q'(t) + \int_{t_0}^{t} \frac{dt'}{t'} \frac{\Delta_q'(t)}{\Delta_q(t')} \Phi_q(t') \frac{C_F}{2\pi} \left[\int_{t_0}^{t'} dt'' \frac{\alpha_s(t'')}{t''} \Phi_g(t'') - \frac{3}{2}\alpha_s(t')\Phi_g(t') \right] \\ &\quad + \frac{1}{t} \frac{\Delta_q(t)}{\Delta_q(t)} \Phi_q(t) \frac{C_F}{2\pi} \left[\int_{t_0}^{t} dt' \frac{\alpha_s(t')}{t'} \Phi_g(t') - \frac{3}{2}\alpha_s(t)\Phi_g(t) \right] \\ &= \frac{\Delta_q'(t)}{\Delta_q(t)} \Phi_q(t) + \Phi_q(t) \left[\frac{1}{t} \int_{t_0}^{t} dt' \frac{\alpha_s(t')}{t'} \Phi_g(t') - \frac{3}{2} \frac{\alpha_s(t)}{t} \Phi_g(t) \right]. \end{aligned} \tag{C.5}$$

The solution to this equation is

$$\begin{aligned}\Phi_q(t) &= u\Delta_q(t)\exp\left[\int_{t_0}^{t} dt' \frac{C_F}{2\pi}\frac{\alpha_s(t')}{t'}\left(\ln\frac{t}{t'}-\frac{3}{2}\right)\Phi_g(t')\right], \\ &= u\Delta_q(t)\exp\left[\int_{t_0}^{t} dt'\Gamma_{q\to qg}(t,t')\Phi_g(t')\right], \end{aligned} \tag{C.6}$$

which we see by differentiation. The computation for $\Phi_g(t)$ is equivalent.

Poisson scaling

If the generating function is of the form

$$\begin{aligned}\Phi(u;t) &= u\exp\left[\int_{t_0}^{t} dt'\Gamma(t,t')\Big(u-1\Big)\right] \\ &= u\Delta_q\exp\left[-u\ln\Delta_q\right], \end{aligned} \tag{C.7}$$

introducing the Sudakov form factor as

$$\Delta_q = \exp\left[-\int_{t_0}^{t} dt'\Gamma(t,t')\right]. \tag{C.8}$$

We prove by induction that

$$\frac{d^n}{du^n}\Phi(u) = \Big(-\ln\Delta_q\Big)^{n-1}\Delta_q\exp\left[-u\ln\Delta_q\right]\Big(n+u\Big(-\ln\Delta_q\Big)\Big). \tag{C.9}$$

Note first that

$$\begin{aligned}\frac{d}{du}\Phi &= \Delta_q\exp\left[-u\ln\Delta_q\right]\Big(1+u\Big(-\ln\Delta_q\Big)\Big) \\ &\text{and} \\ \frac{d^2}{du^2}\Phi &= \Big(-\ln\Delta_q\Big)\Delta_q\exp\left[-u\ln\Delta_q\right]\Big(2+u\Big(-\ln\Delta_q\Big)\Big). \end{aligned} \tag{C.10}$$

Now we compute

$$\begin{aligned}\frac{d^{n+1}}{du^{n+1}}\Phi &= \frac{d}{du}\Big(-\ln\Delta_q\Big)^{n-1}\Delta_q\exp\left[-u\ln\Delta_q\right]\Big(n+u\Big(-\ln\Delta_q\Big)\Big) \\ &= \Big(-\ln\Delta_q\Big)^{n-1}\Delta_q\exp\left[-u\ln\Delta_q\right]\Big(n+u\Big(-\ln\Delta_q\Big)\Big)\Big(-\ln\Delta_q\Big) \\ &\quad + \Big(-\ln\Delta_q\Big)^{n-1}\Delta_q\exp\left[-u\ln\Delta_q\right]\Big(-\ln\Delta_q\Big) \\ &= \Big(-\ln\Delta_q\Big)^{n}\Delta_q\exp\left[-u\ln\Delta_q\right]\Big((n+1)+u\Big(-\ln\Delta_q\Big)\Big). \end{aligned} \tag{C.11}$$

This gives us for the jet rates

$$\begin{aligned}P(n) &= \frac{1}{n!}\frac{d^n}{du^n}\Phi(u)\bigg|_{u=0} \\ &= \frac{\Big(-\ln\Delta_q\Big)^{n-1}}{(n-1)!}\Delta_q, \end{aligned} \tag{C.12}$$

which is a Poisson distribution. The occurrence of $(n-1)$ tells us that for starting with one hard parton, there is no zero jet rate. Thus we reformulate the formula in forms of the probability to radiate additional soft gluons.

$$P_{\text{soft}}(n) = \frac{\lambda^n}{n!}e^{-\lambda}. \tag{C.13}$$

Note, that $n = 0$ now means no *additional* soft radiation. We also use a common factor λ instead of the Sudakov. Depending on the number of hard partons to start with, λ is a combination of quark and gluon Sudakovs. The cross section ratios for Poisson scaling go like

$$R_{\frac{n+1}{n}} = \frac{\lambda}{n+1}. \tag{C.14}$$

We have to be careful though. If we are to use this formula in a fit, the pole is not at $n = -1$, but depends on the number of hard partons. Thus for e^+e^- it is at least at $n = 0$ or even at $n = 1$.

Staircase scaling

If the generating function is described by the differential equation

$$\begin{aligned} \Phi'(t) &= \Gamma(t,t_0)\Phi(t)\Big(\Phi(t)-1\Big) \\ &\text{and} \\ \Phi(t_0) &= u, \end{aligned} \tag{C.15}$$

we can solve it exactly. The solution is

$$\begin{aligned} \Phi(u;t) &= \frac{u}{u+(1-u)\exp\left[\int\limits_{t_0}^{t} dt'\Gamma(t,t_0)\right]} \\ &= \frac{u}{u+(1-u)\tilde{\Delta}_g^{-1}}, \end{aligned} \tag{C.16}$$

introducing the short notion

$$\tilde{\Delta}_g = \exp\left[-\int\limits_{t_0}^{t} dt'\Gamma(t,t_0)\right], \tag{C.17}$$

which is very similar to the Sudakov form factor, but note the different scales in the branching Γ. We prove by differentiation that eq. (C.16) solves the differential equation.

$$\begin{aligned} \frac{d}{dt}\Phi(t) &= \frac{-u}{\left(u+(1-u)\exp\left[\int\limits_{t_0}^{t} dt'\Gamma(t,t_0)\right]\right)^2} \times \left(-u+\ u\ +\ (1-u)\exp\left[\int\limits_{t_0}^{t} dt'\Gamma(t,t_0)\right]\right)\Gamma(t,t_0) \\ &= \Gamma(t,t_0)\Phi(t)\Big(\Phi(t)-1\Big). \end{aligned} \tag{C.18}$$

Now we compute the derivatives with respect to u

$$\begin{aligned} \frac{d}{du}\Phi(u) &= \left\{\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-1} + u(-1)\left(u+(1-u)\tilde{\Delta}_g^{-2}\right)^{-2}\left[1-\tilde{\Delta}_g^{-1}\right]\right\} \\ \frac{d^2}{du^2}\Phi(u) &= \left[1-\tilde{\Delta}_g^{-1}\right]\left\{2*(-1)\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-2} + u(-1)(-2)\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-3}\left[1-\tilde{\Delta}_g^{-1}\right]\right\} \\ \frac{d^3}{du^3}\Phi(u) &= \left[1-\tilde{\Delta}_g^{-1}\right]^2\left\{3*(-1)(-2)\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-3} + u(-1)(-2)(-3)\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-4}\left[1-\tilde{\Delta}_g^{-1}\right]\right\} \\ \frac{d^n}{du^n}\Phi(u) &= n!\left[\tilde{\Delta}_g^{-1}-1\right]^{n-1}\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-n}\left\{1+u\ \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-1}\left[\tilde{\Delta}_g^{-1}-1\right]\right\} \end{aligned} \tag{C.19}$$

That the last equality holds we prove by induction

$$\begin{aligned}
\frac{d^{n+1}}{du^{n+1}}\Phi(u) &= \frac{d}{du} n! \left[\tilde{\Delta}_g^{-1}-1\right]^{n-1} \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-n} \left\{1+u\ \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-1}\left[\tilde{\Delta}_g^{-1}-1\right]\right\} \\
&= n!\left[\tilde{\Delta}_g^{-1}-1\right]^{n-1} \times \\
&\quad \left[n\left[\tilde{\Delta}_g^{-1}-1\right]\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-n-1}\left\{1+u\ \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-1}\left[\tilde{\Delta}_g^{-1}-1\right]\right\}\right. \\
&\quad + \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-n}\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-1}\left[\tilde{\Delta}_g^{-1}-1\right] \\
&\quad \left. + \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-n} u\ \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-2}\left[\tilde{\Delta}_g^{-1}-1\right]^2\right] \\
&= (n+1)!\left[\tilde{\Delta}_g^{-1}-1\right]^{n}\left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-(n+1)}\left\{1+u\ \left(u+(1-u)\tilde{\Delta}_g^{-1}\right)^{-1}\left[\tilde{\Delta}_g^{-1}-1\right]\right\}
\end{aligned} \tag{C.20}$$

Thus we get for the jet rates

$$\begin{aligned}
S(n) &= \left.\frac{1}{n!}\frac{d^n}{du^n}\Phi(u)\right|_{u=0} \\
&= \tilde{\Delta}_g\left[1-\tilde{\Delta}_g\right]^{n-1},
\end{aligned} \tag{C.21}$$

which follows staircase scaling.

C.2 Exponentiated form of the evolution equation for the generalized k_T algorithm

We start from [124]

$$\begin{aligned}
\Phi_q(E,\xi) &= u + \int_{\xi_R}^{\xi}\frac{d\xi'}{\xi'}\int_{E_R/E}^{1} dz \frac{\alpha_s(k_T^2)}{2\pi} P_{q\to qg}\Phi_q(E,\xi')\left[\Phi_g(zE,\xi')-1\right] \\
\Phi_g(E,\xi) &= u + \int_{\xi_R}^{\xi}\frac{d\xi'}{\xi'}\int_{E_R/E}^{1} dz \frac{\alpha_s(k_T^2)}{2\pi} \left\{P_{g\to gg}\Phi_g(E,\xi')\left[\Phi_g(zE,\xi')-1\right]\right. \\
&\qquad \left. + P_{g\to q\bar{q}}\left[\Phi_q^2(E,\xi')-\Phi_g(E,\xi')\right]\right\}.
\end{aligned} \tag{C.22}$$

This we rewrite in the form

$$\Phi_i(e,\xi) = u + \int_{\xi_R}^{\xi}\frac{d\xi'}{\xi'}\int_{1/e}^{1} dz \frac{\alpha_s(z,\xi')}{2\pi}\sum_{j,l}\left\{P_{i\to jl}(z)\Phi_i(e,\xi')\left[\frac{\Phi_j(e,\xi')\Phi_l(E(z),\xi')}{\Phi_i(e,\xi')}-1\right]\right\}. \tag{C.23}$$

i is either q or g and we sum over all allowed splittings. Now we take the derivative of ξ and find

$$\frac{d\Phi_i(E,\xi)}{d\xi} = \frac{1}{\xi}\Phi_i(E,\xi)\int_{E_R/E}^{1} dz \frac{\alpha_s(z,\xi)}{2\pi}\sum_{j,l}\left\{P_{i\to jl}(z)\left[\frac{\Phi_j(E,\xi)\Phi_l(E(z),\xi)}{\Phi_i(E,\xi)}-1\right]\right\}. \tag{C.24}$$

The solution is

$$\Phi_i(E,\xi) = u\exp\left[\int_{\xi_R}^{\xi}\frac{d\xi'}{\xi'}\int_{E_R/E}^{1} dz \frac{\alpha_s(z,\xi')}{2\pi}\sum_{j,l}\left\{P_{i\to jl}(z)\left[\frac{\Phi_j(E,\xi)\Phi_l(E(z),\xi)}{\Phi_i(E,\xi)}-1\right]\right\}\right]. \tag{C.25}$$

C.3 Closed solution in the staircase limit with breaking terms

We start from eq. (1.39)

$$\Phi_g(e,\xi) = u\,\exp\left[\int_{\xi_R}^{\xi}\frac{d\xi'}{\xi'}\int_{1/e}^{1}dz\frac{\alpha_s(z,\xi')}{2\pi}\left\{P_{g\to gg}(z)\left[\left(\Phi_g(e,\xi')+\sum_{n=1}^{\infty}\frac{(e(z-1))^n}{n!}\frac{d^n\Phi_g(e,\xi')}{de^n}\right)-1\right]\right\}\right]. \tag{C.26}$$

Note that due its exponantial form the derivativ of Φ_g can always be written in the following form

$$\frac{d^n\Phi_g(e,\xi')}{de^n} = \Phi_g(e,\xi')\times\mathcal{DP}[n](e,\xi';\Phi_g), \tag{C.27}$$

where $\mathcal{DP}[n]$ is a polynomal of derivatives of Φ_g. We plug this in eq. (C.26) and get

$$\begin{aligned}\Phi_g(e,\xi) = u\,\exp\Bigg[&\int_{\xi_R}^{\xi}\frac{d\xi'}{\xi'}\int_{1/e}^{1}dz\frac{\alpha_s(z,\xi')}{2\pi}\times\\ &P_{g\to gg}(z)\left[\left(\Phi_g(e,\xi')+\Phi_g(e,\xi')\underbrace{\sum_{n=1}^{\infty}\frac{(e(z-1))^n}{n!}\mathcal{DP}[n](e,\xi';\Phi_g)}_{\mathcal{T}(z,e,\xi')}\right)-1\right]\Bigg].\end{aligned} \tag{C.28}$$

Up to this point we have not gained any new insight. All we do is using exact reformulations and introducing new symbols. To evolve further we need to employ some assumptions. This is exactly what we mean by dropping the explicit Φ_g dependence in $\mathcal{T}$. We assume that it is possible, in some approximation of course, to write down such an equation. For example we could expand around $e\approx 1$ and write Φ_g as an explicit series in u, dropping higher terms. At the moment we just assume that such an approximation exists. We can now rewrite eq. (C.28) and bring it in the form

$$\begin{aligned}\Phi_g(e,\xi) = u\,\exp\Bigg[&\int_{\xi_R}^{\xi}d\xi'\left[\Phi_g(e,\xi')-1\right]\underbrace{\frac{1}{\xi'}\int_{1/e}^{1}dz\frac{\alpha_s(z,\xi')}{2\pi}P_{g\to gg}(z)}_{\gamma_g(e,\xi')}+\\ &\int_{\xi_R}^{\xi}d\xi'\Phi_g(e,\xi')\underbrace{\frac{1}{\xi'}\int_{1/e}^{1}dz\frac{\alpha_s(z,\xi')}{2\pi}P_{g\to gg}(z)\mathcal{T}(z,e,\xi')}_{r(e,\xi')}\Bigg]\\ = u\,\exp\Bigg[&\int_{\xi_R}^{\xi}d\xi'\left[\Phi_g(e,\xi')-1\right]\gamma_g(e,\xi')+\int_{\xi_R}^{\xi}d\xi'\Phi_g(e,\xi')r(e,\xi')\Bigg].\end{aligned} \tag{C.29}$$

Taking the derivative this defines a differential equation of the form[1]

$$\begin{aligned}\frac{d\Phi_g(e,\xi)}{d\xi} &= \Phi_g(e,\xi)\times\left[\gamma_g(e,\xi)\left(\Phi_g(e,\xi)-1\right)+r(e,\xi)\Phi_g(e,\xi)\right]\\ \Phi_g(e,\xi_r) &= u.\end{aligned} \tag{C.30}$$

[1] The differential equation itsself is still valid even if we take the Φ_g dependence of $r(e,\xi)$ into account. Nevertheless, we can only solve the equation if we can get rid of it.

We note that for the case $r(e,\xi) \to 0$ Eq. (C.30) produces exact staircase scaling [29]. Thus we connect this term to staircase breaking. We also can solve eq. (C.30) for arbitrary $r(e,\xi)$. The solution is

$$\Phi_g(e,\xi) = \frac{1}{1 + \frac{(1-u)}{u\Delta_g(e,\xi)} - \underbrace{\int_{\xi_R}^{\xi} d\xi' \frac{\Delta_g(e,\xi')}{\Delta_g(e,\xi)} r(e,\xi')}_{\mathcal{R}(e,\xi)}} \tag{C.31}$$

where $\Delta_g(e,\xi) = \exp\left[-\int_{\xi_R}^{\xi} d\xi' \gamma(e,\xi')\right]$ is the Sudakov form factor. We check the solution explicitly by differentiating eq. (C.31) with respect to ξ. We note that we have

$$\begin{aligned} \frac{d\Delta_g(e,\xi)}{d\xi} &= -\gamma(e,\xi)\Delta_g(e,\xi) \\ \frac{d\mathcal{R}(e,\xi)}{d\xi} &= r(e,\xi) + \gamma(e,\xi)\mathcal{R}(e,\xi). \end{aligned} \tag{C.32}$$

Therefor, we get

$$\begin{aligned} \frac{d\Phi_g(e,\xi)}{d\xi} &= (-1)\left(1 + \frac{1-u}{u\Delta_g(e,\xi)} - \mathcal{R}(e,\xi)\right)^{-2} \\ &\quad \times \left[\frac{1-u}{u\Delta_g(e,\xi)}\gamma(e,\xi)\Delta_g(e,\xi) - r(e,\xi) - \gamma(e,\xi)\mathcal{R}(e,\xi)\right] \\ &= \Phi_g(e,\xi) \times \left[\gamma_g(e,\xi)\left(\Phi_g(e,\xi) - 1\right) + r(e,\xi)\Phi_g(e,\xi)\right]. \end{aligned} \tag{C.33}$$

There is one important step left if we actually like to compute jet rates from this equation. We need to know the u dependence of $\mathcal{R}$. Plugging in all definitions we have

$$\begin{aligned} \mathcal{R}(e,\xi) &= \int_{\xi_R}^{\xi} d\xi' \frac{\Delta_g(e,\xi')}{\Delta_g(e,\xi)} r(e,\xi') \\ &= \int_{\xi_R}^{\xi} d\xi' \frac{\Delta_g(e,\xi')}{\Delta_g(e,\xi)} \frac{1}{\xi'} \int_{1/e}^{1} dz \frac{\alpha_s(z,\xi')}{2\pi} P_{g\to gg}(z)\mathcal{T}(z,e,\xi') \\ &= \int_{\xi_R}^{\xi} d\xi' \frac{\Delta_g(e,\xi')}{\Delta_g(e,\xi)} \frac{1}{\xi'} \int_{1/e}^{1} dz \frac{\alpha_s(z,\xi')}{2\pi} P_{g\to gg}(z) \sum_{n=1}^{\infty} \frac{(e(z-1))^n}{n!} \mathcal{DP}[n](e,\xi';\Phi_g) \end{aligned} \tag{C.34}$$

The task at hand is to find a Φ_g independent approximation for $\mathcal{DP}[n]$. An obvious step is to truncate the Tayler expansion at $n = 0$. We are then left with

$$\mathcal{DP}[1](e,\xi';\Phi_g) = \frac{d\Phi_g(e,\xi')}{\Phi_g(e,\xi')de}. \tag{C.35}$$

We need to find the significant part in the limit $e \to 1$ and its u dependence.

$$\begin{aligned} \mathcal{DP}[1](e,\xi';\Phi_g) = \int_{\xi_R}^{\xi'} \frac{d\xi''}{\xi''} \Bigg[&\frac{\alpha_s(1/e,\xi'')P_{g\to gg}(1/e)}{2\pi\ e^2} \left(\underbrace{\Phi_g(1,\xi'')}_{=u} - 1\right) + \\ &\int_{1/e}^{1} dz \frac{\alpha_s(z,\xi'')P_{g\to gg}(z)z}{2\pi} \left[\frac{d\Phi(e,\xi'')}{de}\bigg|_{e=ze}\right] \Bigg] \end{aligned} \tag{C.36}$$

In principle we get a nested series of Φ_g differentiations. To argue that first term is the most important we note that formally in the $e \to 1$ limit we can write

$$\int_{1/e}^{1} dz \frac{\alpha_s(z,\xi'')P_{g\to gg}(z)z}{2\pi}\left[\frac{d\Phi(e,\xi'')}{de}\bigg|_{e=ze}\right] \approx (1-1/e)\frac{\alpha_s(1/e,\xi'')P_{g\to gg}(1/e)}{2\pi\ e}\left[\frac{d\Phi(e,\xi'')}{de}\bigg|_{e\approx 1}\right]. \tag{C.37}$$

We note that the last term of eq. (C.37) is exactly zero in the formal limit $e \to 1$. This is our first argument that the second term vanishes faster than the first one in the $e \to 1$ limit. We also can controll its size this way and we set it to $u-1$ which is for sure over estimated as at the end of the day we have $u \to 0$. We than see that for $e > 1$ and $e \to 1$

$$\begin{aligned} 1 &> e-1 \\ \frac{1}{e} &> \frac{e-1}{e} \\ \frac{1}{e} &> (1-1/e). \end{aligned} \tag{C.38}$$

Therefor, the first term is greater than the second one. However, this is only valid in the $e \to 1$ limit. From this estimate we can not deduce the uncertainty we introduce by dropping the second term nor can we conclude on the range of e where this estimate is valid, because we do not know the correct e dependence. Nevertheless, we compute formally the u dependence of the staircase breaking term in the staircase limit $e \to 1$. Thus we write

$$\begin{aligned} \mathcal{DP}[1](e,\xi';\Phi_g) &\approx \int_{\xi_R}^{\xi'} \frac{d\xi''}{\xi''}\frac{\alpha_s(1/e,\xi'')P_{g\to gg}(1/e)}{2\pi\ e^2}(u-1) \\ &\equiv (u-1)\rho(e,\xi') \\ \mathcal{R}(e,\xi) &= \int_{\xi_R}^{\xi} d\xi' \frac{\Delta_g(e,\xi')}{\Delta_g(e,\xi)}\frac{1}{\xi'}\int_{1/e}^{1} dz \frac{\alpha_s(z,\xi')}{2\pi}P_{g\to gg}(z)(z-1)e(u-1)\rho(e,\xi') \\ &\equiv (u-1)\chi(e,\xi). \end{aligned} \tag{C.39}$$

C.4 Jet rates with i initial soft gluons

In app. C.1 we compute the behavior of one soft gluon initializing the shower cascade. It is very useful to generalize this formula to the case of i initial gluons.

$$\begin{aligned} S(n;1) &= \tilde{\Delta}_g\left[1-\tilde{\Delta}_g\right]^{n-1} \\ S(n;2) &= \sum_{k=1}^{n-1} S(k;1)\times S(n-k;1) \\ &= (n-1)\tilde{\Delta}_g\left[1-\tilde{\Delta}_g^2\right]^{n-2} \\ S(n;3) &= \sum_{k=2}^{n-1} S(k;2)\times S(n-k;1) \\ &= \frac{1}{2}(n-1)(n-2)\tilde{\Delta}_g^3\left[1-\tilde{\Delta}_g\right]^{n-3} \\ S(n;i) &= \frac{(n-1)!}{(n-i)!}\frac{\tilde{\Delta}_g^i}{(i-1)!}\left[1-\tilde{\Delta}_g^2\right]^{n-i}. \end{aligned} \tag{C.40}$$

To show that the last equality is true in general we perform the induction step

$$\begin{aligned}
S(n;i+1) &= \sum_{k=i}^{n-1} S(k;i) \times S(n-k;1) \\
&= \sum_{k=i}^{n-1} \frac{(k-1)!}{(k-i)!} \frac{\tilde{\Delta}_g^i}{(i-1)!} \left[1-\tilde{\Delta}_g^2\right]^{k-i} \times \tilde{\Delta}_g \left[1-\tilde{\Delta}_g\right]^{n-k-1} \\
&= \frac{\tilde{\Delta}_g^{i+1}}{(i-1)!} \left[1-\tilde{\Delta}_g^2\right]^{n-(i+1)} \sum_{k=i}^{n-1} \frac{(k-1)!}{(k-i)!} \\
&= \frac{\tilde{\Delta}_g^{i+1}}{(i-1)!} \left[1-\tilde{\Delta}_g^2\right]^{n-(i+1)} \frac{(n-1)!}{(n-(i+1))!} \frac{1}{i} \\
&= \frac{(n-1)!}{(n-(i+1))!} \frac{\tilde{\Delta}_g^{i+1}}{((i+1)-1)!} \left[1-\tilde{\Delta}_g^2\right]^{n-(i+1)}
\end{aligned} \tag{C.41}$$

Note, that for $n >> i$ the distribution gets staircase like again.

Bibliography

[1] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716**, 1 (2012).

[2] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716**, 30 (2012).

[3] P. W. Higgs, Phys. Lett. **12**, 132 (1964); P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964);

[4] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).

[5] S. Weinberg, Cambridge, UK: Univ. Pr. (1995) 609 p

[6] M. M. Nojiri, T. Plehn, G. Polesello, J. M. Alexander, B. C. Allanach, A. J. Barr, K. Benakli and F. Boudjema *et al.*, arXiv:0802.3672 [hep-ph]. D. E. Morrissey, T. Plehn and T. M. P. Tait, Phys. Rept. **515**, 1 (2012) [arXiv:0912.3259 [hep-ph]].

[7] D. B. Kaplan, H. Georgi and S. Dimopoulos, Phys. Lett. B **136**, 187 (1984).

[8] M. Perelstein, Prog. Part. Nucl. Phys. **58**, 247 (2007) [hep-ph/0512128].

[9] C. Englert, T. Plehn, M. Rauch, D. Zerwas and P. M. Zerwas, Phys. Lett. B **707**, 512 (2012) [arXiv:1112.3007 [hep-ph]].

[10] M. Dine, N. Seiberg and S. Thomas, Phys. Rev. D **76**, 095004 (2007) [arXiv:0707.0005 [hep-ph]]. C. Englert, A. Freitas, M. Muhlleitner, T. Plehn, M. Rauch, M. Spira and K. Walz, arXiv:1403.7191 [hep-ph].

[11] G. Weiglein *et al.* [LHC/LC Study Group Collaboration], Phys. Rept. **426**, 47 (2006); M. Klute, R. Lafaye, T. Plehn, M. Rauch and D. Zerwas, Europhys. Lett. **101**, 51001 (2013).

[12] L. Rossi, Conf. Proc. C **110904**, 908 (2011). A. Rossi, R. W. Assmann and D. Wollmann, Conf. Proc. C **110904**, 3751 (2011). O. S. Bruning, R. Cappi, R. Garoby, O. Grobner, W. Herr, T. Linnecar, R. Ostojic and K. Potter *et al.*, CERN-LHC-PROJECT-REPORT-626.

[13] Aleksa, M (CERN) ; Cleland, W (Pittsburgh) ; Enari, Y (Tokyo) ;Fincke-Keeler, M (Victoria) ; Hervas, L (CERN) ; Lanni, F (BNL) ; Majewski, CERN-LHCC-2013-017 ; ATLAS-TDR-022. - 2013. Capeans, M ; Darbo, G ; Einsweiller, K ; Elsing, M ; Flick, T ; Garcia-Sciveres, M ; Gemme, C ; Pernegger, H ; Rohne, O ; Vuillermet, R CERN-LHCC-2010-013 ; ATLAS-TDR-19. - 2010.

[14] D. Contardo and J. Spalding CMSBUL-ARTICLE-2013-049.- 2013 - Published in : , no. 3/2013, pp. 16

[15] LHCb, Collaboration CERN-LHCC-2014-016 ; LHCB-TDR-016. - 2014.

[16] Antonioli, P (ed.) ; Kluge, A (ed.) ; Riegler, W (ed.) CERN-LHCC-2013-019 ; ALICE-TDR-015. - 2013.

[17] M. Klute, R. Lafaye, T. Plehn, M. Rauch and D. Zerwas, Phys. Rev. Lett. **109**, 101801 (2012) [arXiv:1205.2699 [hep-ph]].

[18] J. S. Kim, K. Rolbiecki, K. Sakurai and J. Tattersall, arXiv:1406.0858 [hep-ph]. M. Drees, H. Dreiner, D. Schmeier, J. Tattersall and J. S. Kim, arXiv:1312.2591 [hep-ph].

[19] G. Aad *et al.* [ATLAS Collaboration], JHEP **1304**, 075 (2013) [arXiv:1210.4491 [hep-ex]]. G. Aad *et al.* [ATLAS Collaboration], arXiv:1407.7494 [hep-ex]. G. Aad *et al.* [ATLAS Collaboration], arXiv:1407.6583 [hep-ex].

[20] S. Chatrchyan *et al.* [CMS Collaboration], Eur. Phys. J. C **73**, 2568 (2013) [arXiv:1303.2985 [hep-ex]]. S. Chatrchyan *et al.* [CMS Collaboration], Phys. Rev. Lett. **109**, 171803 (2012) [arXiv:1207.1898 [hep-ex]].

[21] M. Srednicki, Cambridge, UK: Univ. Pr. (2007) 641 p

[22] M. E. Peskin and D. V. Schroeder, Reading, USA: Addison-Wesley (1995) 842 p

[23] S. Pokorski, Cambridge, Uk: Univ. Pr. (1987) 394 P. (Cambridge Monographs On Mathematical Physics)

[24] T. Plehn, Lect. Notes Phys. **844**, 1 (2012) [arXiv:0910.4182 [hep-ph]].

[25] R. K. Ellis, W. J. Stirling and B. R. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8**, 1 (1996).

[26] P. Schichtel, http://www.thphys.uni-heidelberg.de/ plehn/includes/theses/schichtel.pdf

[27] C. Englert, T. Plehn, P. Schichtel and S. Schumann, Phys. Rev. D **83**, 095009 (2011) [arXiv:1102.4615 [hep-ph]].

[28] C. Englert, T. Plehn, P. Schichtel and S. Schumann, JHEP **1202**, 030 (2012) [arXiv:1108.5473 [hep-ph]].

[29] E. Gerwick, T. Plehn, S. Schumann and P. Schichtel, JHEP **1210**, 162 (2012) [arXiv:1208.3676 [hep-ph]].

[30] E. Gerwick and P. Schichtel, work in progress.

[31] E. Gerwick, T. Plehn and S. Schumann, Phys. Rev. Lett. **108**, 032003 (2012) [arXiv:1108.3335 [hep-ph]].

[32] C. Bernaciak, M. S. A. Buschmann, A. Butter and T. Plehn, Phys. Rev. D **87**, 073014 (2013) [arXiv:1212.4436 [hep-ph]].

[33] C. Bernaciak, B. Mellado, T. Plehn, P. Schichtel and X. Ruan, Phys. Rev. D **89**, 053006 (2014) [arXiv:1311.5891 [hep-ph]].

[34] The ATLAS experiment, http://atlas.ch/

[35] The CMS experiment, http://cms.web.cern.ch/

[36] Particle Data Group, http://pdg.lbl.gov/

[37] The LHCb collaboration, https://lhcb.web.cern.ch/lhcb/

[38] M. Cacciari and G. P. Salam, Phys. Lett. B **641** (2006) 57; M. Cacciari, G. P. Salam and G. Soyez, `http://fastjet.fr`. M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012) [arXiv:1111.6097 [hep-ph]].

[39] G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).

[40] S. Catani and M. H. Seymour, Nucl. Phys. B **485**, 291 (1997) [Erratum-ibid. B **510**, 503 (1998)] [hep-ph/9605323]. S. Catani, S. Dittmaier and Z. Trocsanyi, Phys. Lett. B **500**, 149 (2001) [hep-ph/0011222]. S. Catani, S. Dittmaier, M. H. Seymour and Z. Trocsanyi, Nucl. Phys. B **627**, 189 (2002) [hep-ph/0201036].

[41] Dorival Gonzales-Netto, http://www.thphys.uni-heidelberg.de/ plehn/includes/theses/netto.pdf

[42] I. O. Cherednikov and N. G. Stefanis, Nucl. Phys. B **802**, 146 (2008) [arXiv:0802.2821 [hep-ph]]. I. O. Cherednikov and N. G. Stefanis, Phys. Rev. D **77**, 094001 (2008) [arXiv:0710.1955 [hep-ph]]. L. D. McLerran and R. Venugopalan, Phys. Rev. D **49**, 3352 (1994) [hep-ph/9311205].

[43] Anna Kulesza, George Sterman, and Werner Vogelsang, Phys. Rev. D **69**, 014012 (2004).

[44] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, JHEP **1010**, 078 (2010) [arXiv:1006.2833 [hep-ph]]. T. Plehn and M. Spannowsky, J. Phys. G **39**, 083001 (2012) [arXiv:1112.4441 [hep-ph]].

[45] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **0111**, 063 (2001) [hep-ph/0109231]. F. Krauss, JHEP **0208** (2002) 015. S. Höche, F. Krauss, S. Schumann and F. Siegert, JHEP **0905**, 053 (2009).

[46] ML. Mangano, 2004 - cepa.fnal.gov M. L. Mangano, M. Moretti and R. Pittau, Nucl. Phys. B **632** (2002) 343 .

[47] T. Gleisberg, S. .Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, JHEP **0902**, 007 (2009) [arXiv:0811.4622 [hep-ph]]. S. Schumann, F. Krauss, JHEP **0803**, 038 (2008); T. Gleisberg, S. Höche, JHEP **0812**, 039 (2008).

[48] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, JHEP **1106**, 128 (2011) [arXiv:1106.0522 [hep-ph]]. J. Alwall, P. Demin, S. de Visscher *et al.*, JHEP **0709** (2007) 028.

[49] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, JHEP **0307**, 001 (2003) [hep-ph/0206293].

[50] T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008) [arXiv:0710.3820 [hep-ph]]. T. Sjostrand, S. Mrenna, P. Z. Skands, JHEP **0605** (2006) 026.

[51] M. Bahr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer and P. Richardson *et al.*, Eur. Phys. J. C **58**, 639 (2008) [arXiv:0803.0883 [hep-ph]]. G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour and B. R. Webber, JHEP **0101**, 010 (2001);

[52] D. Krohn, J. Thaler and L. -T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].

[53] D. E. Morrissey, T. Plehn and T. M. P. Tait, arXiv:0912.3259 [hep-ph]. A. Djouadi, Phys. Rept. **457**, 1 (2008); T. Plehn, Lect. Notes Phys. **844**, 1 (2012) [arXiv:0910.4182 [hep-ph]].

[54] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716**, 1 (2012). ATLAS Collaboration, ATLAS-CONF-2010-065.

[55] CMS Collaboration, arXiv:1101.1628 [hep-ex]. S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716**, 30 (2012).

[56] D. S. M. Alves, E. Izaguirre and J. G. Wacker, arXiv:1008.0407 [hep-ph].

[57] H. Baer, C. Chen, F. Paige *et al.*, Phys. Rev. **D52** (1995) 2746-2759. I. Hinchliffe, F. E. Paige, M. D. Shapiro *et al.*, Phys. Rev. **D55** (1997) 5520-5540; L. Randall, D. Tucker-Smith, Phys. Rev. Lett. **101** (2008) 221803; A. J. Barr, C. Gwenlan, Phys. Rev. **D80** (2009) 074007.

[58] G. Aad *et al.* [ATLAS Collaboration], JINST **3** (2008) S08003.

[59] G. L. Bayatian *et al.* [CMS Collaboration], J. Phys. G **34** (2007) 995.

[60] for an early correct description of one jet over its entire p_T range see: M. Bengtsson and T. Sjostrand, Phys. Lett. B **185** (1987) 435; M. Bengtsson and T. Sjostrand, Nucl. Phys. B **289** (1987) 810; G. Miu and T. Sjostrand, Phys. Lett. B **449** (1999) 313; E. Norrbin and T. Sjostrand, Nucl. Phys. B **603** (2001) 297; M. H. Seymour, Nucl. Phys. **B436** (1995) 443-460; M. H. Seymour, Comput. Phys. Commun. **90** (1995) 95-101; G. Corcella, M. H. Seymour, Nucl. Phys. **B565** (2000) 227-244.

[61] S. Höche, F. Krauss, S. Schumann and F. Siegert, JHEP **0905** (2009) 053.

[62] A. Buckley, J. Butterworth, S. Gieseke *et al.*, arXiv:1101.2599 [hep-ph].

[63] J. Alwall, S. Höche, F. Krauss *et al.*, Eur. Phys. J. **C53** (2008) 473-500.

[64] S. D. Ellis, R. Kleiss, W. J. Stirling, Phys. Lett. **B154** (1985) 435.

[65] V. M. Abazov *et al.* [D0 Collaboration], Phys. Lett. **B658** (2008) 112-119; T. Aaltonen *et al.* [CDF - Run II Collaboration], Phys. Rev. Lett. **100** (2008) 102001.

[66] ATLAS Collaboration, arXiv:1012.5382 [hep-ex].

[67] for a review see e.g. A. J. Barr, C. G. Lester, J. Phys. G **G37** (2010) 123001.

[68] M. Cacciari, G. P. Salam, G. Soyez, JHEP **0804** (2008) 063.

[69] F. A. Berends, W. T. Giele, H. Kuijf, R. Kleiss and W. J. Stirling, Phys. Lett. B **224** (1989) 237; F. A. Berends, H. Kuijf, B. Tausk *et al.*, Nucl. Phys. **B357** (1991) 32.

[70] C. F. Berger, Z. Bern, L. J. Dixon *et al.*, arXiv:1009.2338 [hep-ph].

[71] H. L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin and C. P. Yuan, Phys. Rev. D **82** (2010) 074024.

[72] F. Krauss, A. Schälicke, S. Schumann *et al.*, Phys. Rev. **D70** (2004) 114009; V. M. Abazov *et al.* [DØ Collaboration], Phys. Lett. **B693** (2010) 522-530; S. Höche, S. Schumann, F. Siegert, Phys. Rev. **D81** (2010) 034026; G. Aad *et al.* [ATLAS Collaboration], arXiv:1012.5382 [hep-ex]; DØ Collaboration, Note 6032-CONF; F. A. Dias, E. Nurse and G. Hesketh, arXiv:1102.0917 [hep-ph]; ATLAS Collaboration, arXiv:1102.2696 [hep-ex].

[73] ATLAS Collaboration, ATLAS-CONF-2010-084.

[74] T. Plehn, D. Rainwater, P. Z. Skands, Phys. Lett. **B645** (2007) 217-221.

[75] J. Alwall, S. de Visscher, F. Maltoni, JHEP **0902** (2009) 017.

[76] T. Plehn and T. M. P. Tait, J. Phys. G **36** (2009) 075001.

[77] M. Bahr, S. Gieseke, M. A. Gigg *et al.*, Eur. Phys. J. **C58** (2008) 639-707.

[78] W. Beenakker, R. Höpker, M. Spira and P. M. Zerwas, Nucl. Phys. B **492** (1997) 51; T. Plehn, arXiv:hep-ph/9809319; `www.thphys.uni-heidelberg.de/~tplehn/prospino`

[79] B. C. Allanach, Comput. Phys. Commun. **143** (2002) 305-331.

[80] P. Z. Skands, B. C. Allanach, H. Baer *et al.*, JHEP **0407** (2004) 036.

[81] M. Mühlleitner, A. Djouadi, Y. Mambrini, Comput. Phys. Commun. **168** (2005) 46-70.

[82] B. C. Allanach *et al.*, Eur. Phys. J. C **25** (2002) 113.

[83] J. Alwall, M. -P. Le, M. Lisanti, J. G. Wacker, Phys. Rev. **D79** (2009) 015005.

[84] C. Englert, T. Plehn, P. Schichtel, and S. Schumann, in preparation.

[85] R. Barate *et al.* [LEP Working Group for Higgs boson searches], Phys. Lett. B **565** (2003) 61.

[86] T. Junk, Nucl. Instrum. Meth. A **434** (1999) 435, CDF Note 8128 [cdf/doc/statistics/public/8128], CDF Note 7904 [cdf/doc/statistics/public/7904].

[87] H. Hu and J. Nielsen, in '1st Workshop on Confidence Limits', CERN 2000-005 (2000) [arXiv:physics/9906010]; K. Cranmer and T. Plehn, Eur. Phys. J. C **51** (2007) 415; A. De Rujula, J. Lykken, M. Pierini *et al.*, Phys. Rev. **D82** (2010) 013003; Y. Gao, A. V. Gritsan, Z. Guo *et al.*, Phys. Rev. **D81** (2010) 075022; D. E. Soper and M. Spannowsky, JHEP **1008** (2010) 029.

[88] J. Neyman and E. S. Pearson, Philosophical Transactions of the Royal Society of London. Series A Vol. 231, (1933), pp. 289-337.

[89] E. Gerwick, T. Plehn and S. Schumann, Phys. Rev. Lett. **108**, 032003 (2012).

[90] G. Aad *et al.* [ATLAS Collaboration], Eur. Phys. J. C **72**, 2043 (2012) and arXiv:1207.5644 [hep-ex].

[91] F. Kling, T. Plehn and M. Takeuchi, arXiv:1207.4787 [hep-ph].

[92] C. Englert, T. Plehn, P. Schichtel and S. Schumann, Phys. Rev. D **83**, 095009 (2011).

[93] see e.g. T. Plehn and T. M. P. Tait, J. Phys. G G **36**, 075001 (2009); S. Schumann, A. Renaud and D. Zerwas, JHEP **1109**, 074 (2011); H. K. Dreiner and M. Krämer, and J. Tattersall, arXiv:1207.1613 [hep-ph].

[94] R. M. Duran Delgado, J. R. Forshaw, S. Marzani and M. H. Seymour, JHEP **1108**, 157 (2011) [arXiv:1107.2084 [hep-ph]].

[95] S. D. Ellis, R. Kleiss, W. J. Stirling, Phys. Lett. **B154**, 435 (1985); F. A. Berends, W. T. Giele, H. Kuijf, R. Kleiss, W. J. Stirling, Phys. Lett. **B224**, 237 (1989); F. A. Berends, H. Kuijf, B. Tausk, W. T. Giele, Nucl. Phys. **B357**, 32-64 (1991).

[96] [ATLAS Collaboration], ATLAS-CONF-2011-001.

[97] C. Englert, T. Plehn, P. Schichtel and S. Schumann, JHEP **1202**, 030 (2012).

[98] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller and S. I. Troian, Gif-sur-Yvette, France: Ed. Frontieres (1991) 274 p.

[99] R. K. Ellis, W. J. Stirling and B. R. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8**, 1 (1996).

[100] for a pedagogical introduction see e.g. T. Plehn, Lect. Notes Phys. **844**, 1 (2012) [arXiv:0910.4182 [hep-ph]].

[101] S. Catani, Y. L. Dokshitzer and B. R. Webber, Phys. Lett. B **285**, 291 (1992); S. Catani, B. Webber and Y. L. Dokshitzer, Nucl. Phys. Proc. Suppl. **29A**, 136 (1992).

[102] T. Gleisberg and S. Hoeche, JHEP **0812**, 039 (2008).

[103] T. Becher and M. Neubert, JHEP **1207**, 108 (2012). T. Becher, M. Neubert and L. Rothen, JHEP **1310**, 125 (2013).

[104] A. Banfi, G. P. Salam and G. Zanderighi, JHEP **1206**, 159 (2012).

[105] Z. Bern, G. Diana, L. J. Dixon, F. Febres Cordero, S. Hoeche, D. A. Kosower, H. Ita and D. Maitre *et al.*, arXiv:1112.3940 [hep-ph].

[106] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita and D. A. Kosower *et al.*, Phys. Rev. Lett. **106**, 092001 (2011).

[107] H. Ita, Z. Bern, L. J. Dixon, F. Febres Cordero, D. A. Kosower and D. Maitre, Phys. Rev. D **85**, 031501 (2012).

[108] S. Dittmaier, P. Uwer and S. Weinzierl, Phys. Rev. Lett. **98**, 262002 (2007); A. Bredenstein, A. Denner, S. Dittmaier and S. Pozzorini, JHEP **1003**, 021 (2010); G. Bevilacqua, M. Czakon, C. G. Papadopoulos and M. Worek, Phys. Rev. Lett. **104**, 162002 (2010).

[109] J. M. Campbell, R. K. Ellis and G. Zanderighi, JHEP **0610**, 028 (2006); J. M. Campbell, R. K. Ellis and C. Williams, Phys. Rev. D **81**, 074023 (2010).

[110] G. Ossola, C. G. Papadopoulos and R. Pittau, Nucl. Phys. B **763**, 147 (2007); R. K. Ellis, W. T. Giele and Z. Kunszt, JHEP **0803**, 003 (2008); G. Ossola, C. G. Papadopoulos and R. Pittau, JHEP **0805**, 004 (2008); C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, H. Ita, D. A. Kosower and D. Maitre, Phys. Rev. D **78**, 036003 (2008); R. K. Ellis, W. T. Giele, Z. Kunszt and K. Melnikov, Nucl. Phys. B **822**, 270 (2009); T. Binoth, F. Boudjema, G. Dissertori, A. Lazopoulos, A. Denner, S. Dittmaier, R. Frederix and N. Greiner *et al.*, Comput. Phys. Commun. **181**, 1612 (2010); V. Hirschi, R. Frederix, S. Frixione, M. V. Garzelli, F. Maltoni and R. Pittau, JHEP **1105**, 044 (2011); G. Bevilacqua, M. Czakon, M. V. Garzelli, A. van Hameren, A. Kardos, C. G. Papadopoulos, R. Pittau and M. Worek, arXiv:1110.1499 [hep-ph]; F. Cascioli, P. Maierhofer and S. Pozzorini, Phys. Rev. Lett. **108**, 111601 (2012); G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola, T. Reiter and F. Tramontano, Eur. Phys. J. C **72**, 1889 (2012).

[111] T. Binoth, D. Goncalves Netto, D. Lopez-Val, K. Mawatari, T. Plehn and I. Wigmore, Phys. Rev. D **84**, 075005 (2011); D. Goncalves-Netto, D. Lopez-Val, K. Mawatari, T. Plehn and I. Wigmore, arXiv:1203.6358 [hep-ph].

[112] R. Boughezal, K. Melnikov and F. Petriello, Phys. Rev. D **85**, 034025 (2012); T. Gehrmann and P. F. Monni, JHEP **1112**, 049 (2011); P. Baernreuther, M. Czakon and A. Mitov, arXiv:1204.5201 [hep-ph].

[113] R. Bonciani, S. Catani, M. L. Mangano and P. Nason, Phys. Lett. B **575**, 268 (2003).

[114] A. Banfi, G. P. Salam and G. Zanderighi, JHEP **0503**, 073 (2005).

[115] A. Banfi, P. F. Monni, G. P. Salam and G. Zanderighi, arXiv:1206.4998 [hep-ph].

[116] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, H. Hoeth, F. Krauss and L. Lonnblad *et al.*, Phys. Rept. **504**, 145 (2011).

[117] L. Lonnblad, JHEP **0205**, 046 (2002); L. Lonnblad and S. Prestel, JHEP **1203**, 019 (2012).

[118] J. Alwall, S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, F. Maltoni, M. L. Mangano and M. Moretti *et al.*, Eur. Phys. J. C **53**, 473 (2008).

[119] S. Frixione and B. R. Webber, JHEP **0206**, 029 (2002); S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, arXiv:1111.1220 [hep-ph].

[120] P. Nason, JHEP **0411**, 040 (2004); S. Frixione, P. Nason and C. Oleari, JHEP **0711**, 070 (2007); S. Hoche, F. Krauss, M. Schonherr and F. Siegert, JHEP **1104**, 024 (2011).

[121] K. Hamilton and P. Nason, JHEP **1006**, 039 (2010); S. Hoche, F. Krauss, M. Schonherr and F. Siegert, JHEP **1108**, 123 (2011).

[122] N. Lavesson and L. Lonnblad, JHEP **0812**, 070 (2008); S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, arXiv:1207.5030 [hep-ph]; T. Gehrmann, S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, arXiv:1207.5031 [hep-ph].

[123] G. Aad *et al.* [ATLAS Collaboration], JHEP **1307**, 032 (2013) [arXiv:1304.7098 [hep-ex]].

[124] E. Gerwick, S. Schumann, B. Gripaios and B. Webber, JHEP **1304**, 089 (2013) [arXiv:1212.5235].

[125] D. L. Rainwater, D. Zeppenfeld and K. Hagiwara, Phys. Rev. D **59**, 014037 (1999); T. Plehn, D. L. Rainwater and D. Zeppenfeld, Phys. Rev. D **61**, 093005 (2000); S. Asai *et al.*, Eur. Phys. J. C **32S2**, 19 (2004).

[126] U. Baur, E. W. N. Glover, Phys. Lett. **B252**, 683-689 (1990); V. D. Barger, K. Cheung, T. Han, D. Zeppenfeld, Phys. Rev. **D44**, 2701 (1991); V. D. Barger, R. J. N. Phillips, D. Zeppenfeld, Phys. Lett. **B346**, 106-114 (1995); D. L. Rainwater, R. Szalapski, D. Zeppenfeld, Phys. Rev. **D54**, 6680-6689 (1996); D. L. Rainwater, D. Summers, D. Zeppenfeld, Phys. Rev. **D55**, 5681-5684 (1997).

[127] E. Gerwick, T. Plehn and S. Schumann, arXiv:1108.3335 [hep-ph].

[128] S. Ask, J. H. Collins, J. R. Forshaw, K. Joshi and A. D. Pilkington, arXiv:1108.2396 [hep-ph].

[129] T. Plehn, D. Rainwater, P. Z. Skands, Phys. Lett. **B645** (2007) 217-221; T. Plehn and T. M. P. Tait, J. Phys. G **36** (2009) 075001; J. Alwall, S. de Visscher, F. Maltoni, JHEP **0902** (2009) 017.

[130] Atlas Collaboration, arXiv:1107.2092 [hep-ex].

[131] S. D. Ellis, R. Kleiss, W. J. Stirling, Phys. Lett. **B154**, 435 (1985).

[132] F. A. Berends, W. T. Giele, H. Kuijf, R. Kleiss, W. J. Stirling, Phys. Lett. **B224**, 237 (1989); F. A. Berends, H. Kuijf, B. Tausk, W. T. Giele, Nucl. Phys. **B357**, 32-64 (1991).

[133] The CMS Collaboration, arXiv:1110.3226 [hep-ex].

[134] The CMS Collaboration, CMS PAS EWK-08-006; ATLAS Collaboration, arXiv:1110.2299 [hep-ex].

[135] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438 [Yad. Fiz. **15** (1972) 781]; G. Altarelli and G. Parisi, Nucl. Phys. B **126** (1977) 298; Y. L. Dokshitzer, Sov. Phys. JETP **46** (1977) 641 [Zh. Eksp. Teor. Fiz. **73** (1977) 1216].

[136] for some recent theoretical developments see e.g. I. W. Stewart, F. J. Tackmann, arXiv:1107.2117 [hep-ph]. I. W. Stewart and F. J. Tackmann, Phys. Rev. D **85**, 034011 (2012); I. W. Stewart, F. J. Tackmann, J. R. Walsh and S. Zuberi, arXiv:1307.1808 [hep-ph].

[137] see e.g. J. R. Andersen, L. Lonnblad, J. M. Smillie, JHEP **1107**, 110 (2011); J. Forshaw, J. Keates and S. Marzani, JHEP **0907**, 023 (2009); R. M. D. Delgado, J. R. Forshaw, S. Marzani, M. H. Seymour, arXiv:1107.2084 [hep-ph].

[138] ATLAS Collaboration, arXiv:1012.5382 [hep-ex], arXiv:1106.2061 [hep-ex], arXiv:1107.1641 [hep-ex], and arXiv:1107.2092 [hep-ex]; CMS Collaboration, arXiv:1106.0647 [hep-ex], and arXiv:1110.3226 [hep-ex].

[139] V. M. Abazov *et al.* [D0 Collaboration], Phys. Lett. **B658** (2008) 112-119; T. Aaltonen *et al.* [CDF - Run II Collaboration], Phys. Rev. Lett. **100** (2008) 102001.

[140] M. Escalier, F. Derue, L. Fayard *et al.* CERN-ATL-PHYS-PUB-2005-018; U. Baur, T. Han and J. Ohnemus, Phys. Rev. D **48**, 5140 (1993); S. Frixione, Phys. Lett. B **429** (1998) 369.

[141] S. Höche, S. Schumann, F. Siegert, Phys. Rev. **D81**, 034026 (2010); F. Campanario, C. Englert and M. Spannowsky, Phys. Rev. D **83** (2011) 074009; J. M. Campbell, R. K. Ellis and C. Williams, JHEP **1107**, 018 (2011).

[142] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock and B. R. Webber, Phys. Lett. B **269** (1991) 432.

[143] B. R. Webber, arXiv:1009.5871 [hep-ph].

[144] CMS Collaboration, CMS PAS SUS-08-002; Z. Bern *et al.*, arXiv:1106.1423 [hep-ph]; S. Ask, M. A. Parker, T. Sandoval, M. E. Shea and W. J. Stirling, arXiv:1107.2803 [hep-ph].

[145] P. W. Higgs, Phys. Lett. **12**, 132 (1964); P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964); F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).

[146] G. Weiglein *et al.* [LHC/LC Study Group Collaboration], Phys. Rept. **426**, 47 (2006); M. Klute, R. Lafaye, T. Plehn, M. Rauch and D. Zerwas, Europhys. Lett. **101**, 51001 (2013).

[147] see e.g. M. Shaposhnikov and C. Wetterich, Phys. Lett. B **683**, 196 (2010); M. Holthausen, K. S. Lim and M. Lindner, JHEP **1202**, 037 (2012); A. Hebecker, A. K. Knochel and T. Weigand, Nucl. Phys. B **874**, 1 (2013) D. Buttazzo *et al* arXiv:1307.3536 [hep-ph];

[148] for up-to-date coupling analyses see [ATLAS Collaboration], ATLAS-CONF-2013-034; [ATLAS Collaboration], ATLAS-CONF-2013-040; [CMS Collaboration], CMS-PAS-HIG-13-005; T. Plehn and M. Rauch, Europhys. Lett. **100**, 11002 (2012); D. Lopez-Val, T. Plehn and M. Rauch, JHEP **1310**, 134 (2013); T. Corbett, O. J. P. Eboli, J. Gonzalez-Fraile and M. C. Gonzalez-Garcia, arXiv:1306.0006 [hep-ph]; I. Brivio *et al* arXiv:1311.1823 [hep-ph]. A. Azatov, R. Contino and J. Galloway, JHEP **1204**, 127 (2012); J. Ellis and T. You, JHEP **1306**, 103 (2013); G. F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, JHEP **0706**, 045 (2007); J. R. Espinosa, C. Grojean, M. Mühlleitner and M. Trott, JHEP **1212**, 045 (2012); S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya, JHEP **1210**, 062 (2012); N. Craig and S. Thomas, JHEP **1211**, 083 (2012); F. Bonnet, T. Ota, M. Rauch and W. Winter, Phys. Rev. D **86**, 093014 (2012); A. Djouadi, arXiv:1208.3436 [hep-ph]; B. A. Dobrescu and J. D. Lykken, JHEP **1302**, 073 (2013); E. Massó and V. Sanz, Phys. Rev. D **87**, 033001 (2013); G. Belanger, B. Dumont, U. Ellwanger, J. F. Gunion and S. Kraml, JHEP **1302**, 053 (2013); P. P. Giardino, K. Kannike, I. Masina, M. Raidal and A. Strumia, arXiv:1303.3570 [hep-ph]; A. Djouadi and G. Moreau, arXiv:1303.6591 [hep-ph]; D. Carmi, A. Falkowski, E. Kuflik, T. Volansky and J. Zupan, JHEP **1210**, 196 (2012); P. P. Giardino, K. Kannike, I. Masina, M. Raidal and A. Strumia, arXiv:1303.3570 [hep-ph];

[149] D. L. Rainwater, D. Zeppenfeld and K. Hagiwara, Phys. Rev. D **59**, 014037 (1999); T. Plehn, D. L. Rainwater and D. Zeppenfeld, Phys. Rev. D **61**, 093005 (2000).

[150] N. Kauer, T. Plehn, D. Rainwater and D. Zeppenfeld, Phys. Lett. B **503**, 113 (2001).

[151] D. L. Rainwater and D. Zeppenfeld, JHEP **9712**, 005 (1997); J. R. Andersen, C. Englert and M. Spannowsky, arXiv:1211.3011 [hep-ph].

[152] S. Asai, G. Azuelos, C. Buttar, V. Cavasinni, D. Costanzo, K. Cranmer, R. Harper and K. Jakobs *et al.*, Eur. Phys. J. C **32S2**, 19 (2004).

[153] R. Kleiss and W. J. Stirling, Phys. Lett. B **200**, 193 (1988); U. Baur and E. W. N. Glover, Phys. Lett. B **252**, 683 (1990); V. D. Barger, K. Cheung, T. Han, J. Ohnemus and D. Zeppenfeld, Phys. Rev. D **44**, 1426 (1991).

[154] E. Gerwick, T. Plehn and S. Schumann, Phys. Rev. Lett. **108**, 032003 (2012); E. Gerwick, T. Plehn, S. Schumann and P. Schichtel, JHEP **1210**, 162 (2012); E. Gerwick, S. Schumann, B. Gripaios and B. Webber, JHEP **1304**, 089 (2013).

[155] B. E. Cox, J. R. Forshaw and A. D. Pilkington, Phys. Lett. B **696**, 87 (2011).

[156] D. L. Rainwater, R. Szalapski and D. Zeppenfeld, Phys. Rev. D **54**, 6680 (1996).

[157] A. Banfi, G. P. Salam and G. Zanderighi, JHEP **1206**, 159 (2012) A. Banfi, P. F. Monni, G. P. Salam and G. Zanderighi, Phys. Rev. Lett. **109**, 202001 (2012).

[158] G. C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978); R. D. Field, Y. Kanev and M. Tayebnejad, Phys. Rev. D **55**, 5685 (1997).

[159] C. Bernaciak, M. S. A. Buschmann, A. Butter and T. Plehn, Phys. Rev. D **87**, 073014 (2013).

[160] T. Plehn, D. L. Rainwater and D. Zeppenfeld, Phys. Rev. Lett. **88**, 051801 (2002); C. Ruwiedel, N. Wermes and M. Schumacher, Eur. Phys. J. C **51**, 385 (2007); K. Hagiwara, Q. Li and K. Mawatari, JHEP **0907**, 101 (2009); C. Englert, D. Goncalves-Netto, K. Mawatari and T. Plehn, JHEP **1301**, 148 (2013); A. Djouadi, R. M. Godbole, B. Mellado and K. Mohan, Phys. Lett. B **723**, 307 (2013).

[161] for a nice overview of LHC applications see e.g. A. Banfi, G. P. Salam and G. Zanderighi, JHEP **1006**, 038 (2010).

[162] C. Englert, M. Spannowsky and M. Takeuchi, JHEP **1206**, 108 (2012); C. Englert, D. Goncalves, G. Nail and M. Spannowsky, Phys. Rev. D **88**, 013016 (2013).

[163] A. Höcker *et al.*, PoS ACAT , 040 (2007) [physics/0703039 [PHYSICS]]; P. Speckmayer, A. Höcker, J. Stelzer and H. Voss, J. Phys. Conf. Ser. **219**, 032057 (2010); `http://tmva.sourceforge.net`

[164] [ATLAS Collaboration], ATLAS-CONF-2013-012.

[165] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA (1992)

[166] J. D. Bjorken and S. D. Drell, ISBN-0070054940.

Statement of Authorship

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den

Peter Schichtel