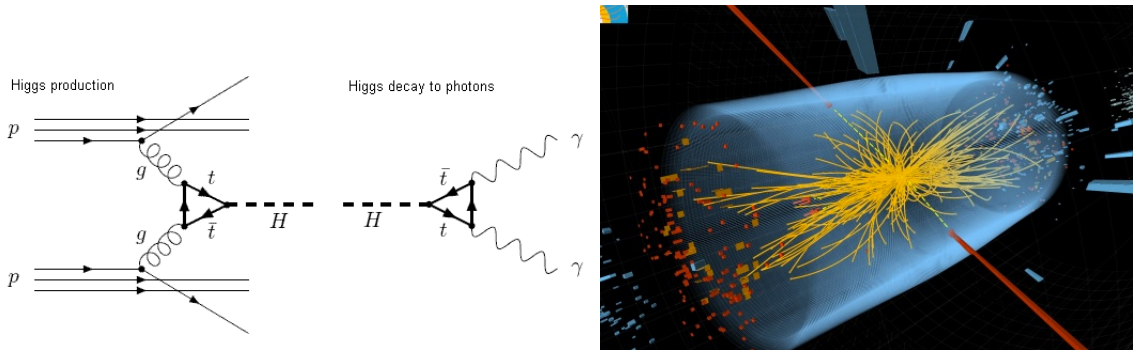


Figure 5.9. Left plot: Feynman diagram for the Higgs boson production in the gg channel, followed by a decay into di-photons. At the partonic level, the event displays a low multiplicity. Right plot: a candidate CMS event for the same process. The number of final state particles (event multiplicity) is much larger.

5.4 Parton showers and Monte Carlo event generators

Up to now, we have discussed **fixed order perturbative QCD calculations**, where the final state has a relatively low **multiplicity** (the number of particles in the final state). See for example the Feynman diagram for Higgs production in gluon channel followed by a decay to di-photons in Fig. 5.9, where only a few particles are seen in the final state of the collision. However, real events look much more complex, as the CMS display for a Higgs candidate event, for the same underlying partonic process, shows.

The connection between the fixed-order QCD calculations in various processes, from e^+e^- annihilation to hadron collisions, and a realistic simulation of the hadronic final state, is achieved by the **Monte Carlo parton shower event generators**. These generators also take care of modeling other aspects of the scattering process, such as **hadronization** (the transition between quark/gluon to hadronic degrees of freedom) which fall outside the regime of perturbative QCD.

Discussion #5.6

Why do we need special tools to generate high-multiplicity events and we cannot just instead compute the **perturbative corrections** up to very high order to the underlying processes, as we have done so far up to NLO? And why are perturbative calculation unable to describe the **parton to hadron transition**, that is, the transition between the quark and gluon degrees of freedom relevant for the hard scattering and the hadronic degrees of freedom that we can detect in our experiments?

Fig. 5.10 provide two schematic representation of the various ingredients that enter the description of a realistic hadronic collision. All these ingredients need be described, either using models or first-principle calculations, in order to compare with realistic experimental measurements. In particular, we need to consider:

(a) First of all, the **initial state**: the probabilities of finding quarks and gluons in the proton carrying a given momentum fraction $f_i(x)$ is described by the PDFs.

(b) Then the **hard scattering reaction** between the partonic constituents, which can be evaluated in perturbative QFT.

(c) Partons with large virtuality $Q^2 \gg \Lambda^2_{\rm QCD} \sim (1\ {\rm GeV})^2$ will lose energy by means of **QCD radiation**, also known as **parton shower**, until they reach virtualities of the order of $Q^2 \sim \Lambda^2_{\rm QCD} \sim (1\ {\rm GeV})^2$. At this point, the perturbative description of the collision stops being valid, because $\alpha_s(Q^2) \sim 1$ and the perturbative expansion breaks down.

(d) This indicates the point at which we need to move from partonic degrees of freedom to hadronic degrees of freedom: **hadronisation**. For this we don't have a good theory framework so we need to rely on models.

(e) Subsequently, these hadrons will **decay** into other particles, and this decay chain is also modeled by the event generator.

(f) In addition to this main processes, other effects can take place during the same proton-proton collision which could influence the final result of the theoretical predictions, such as

- **Multiple parton interactions**: in the same proton-proton collisions, we may have more than one pair of partons that undergo a hard scattering.
- **Pile up**: if the proton bunches crossing in the accelerator have a large multiplicity, it could be that one has more than one proton-proton interaction which then overlap among them.

Theoretical frameworks that hence enable the comprehensive simulation of these various components are hence indispensable for LHC physics.

Discussion #5.7

When designing possible observables in proton-proton collisions, it is important to design them in a way in which they are most sensitive to aspects of the **theory that we have under good control** (PDFs, hard-scattering matrix elements) and less so to other aspects (hadronisation, pile up) for which uncertainties are larger. Taking into account these considerations, discuss which observables in proton-proton collisions can be **evaluated more reliably**:

- Jet production cross-section at low-p_T, say 10 GeV or at high-p_T, say 5 TeV?
- Top quark production in the hadronic final state $t \to W^+ b \to q\bar{q}'b$or in the leptonic final state $t \to W^+ b \to \ell \bar{\nu}_\ell b$?
- The production of two pions $p + p \to \pi^+ + \pi^- + X$ or the production of two jets $p + p \to \text{jet} + \text{jet} + X$?

The **underlying strategy** of Monte Carlo event generators is the following:

- The starting point is the partonic configuration (in terms of quark and gluons) after the hard scattering, when all particles have virtualities $Q^2 \gg \Lambda^2_{\rm qcd}$. These partonic configurations are sampled from the partonic matrix element, which can be evaluated in perturbation theory.
- Then a **parton shower** follows, where quarks an gluons radiate more quarks and gluons via soft and collinear splittings (which are dominant in the matrix elements, and resummed to all orders) until virtualities of all partons in the event is $Q_0^2 \simeq \Lambda^2_{\rm qcd}$.
- The hadronization of colored hadrons into color singlet hadrons is performed using models.
- Finally, **soft and semi-hard components** of the process such as underlying event or multiple parton scattering are also modeled, as well as related effects such as pileup.

The end result of this simulation is a realistic description of the final state in hadronic collisions. Monte Carlo generators are hence based on a combination of first-principle calculations and of phenomenological models, in the latter case with parameters that need to be adjusted to experimental data.

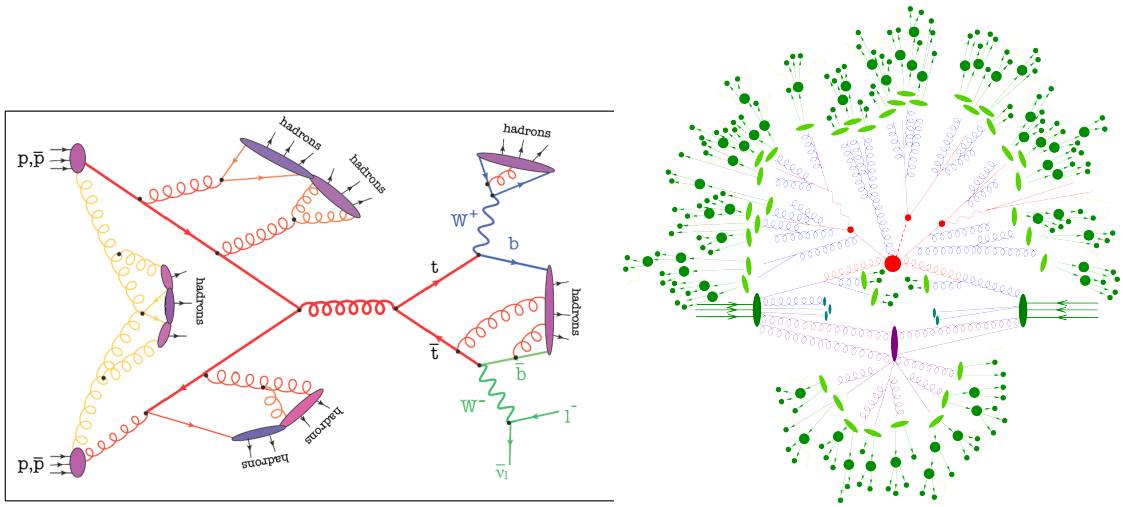


Figure 5.10. Two schematic representation of the various ingredients that enter an hadronic collision, from the initial state (parton distributions) to the hard scattering cross-section to the final state (parton shower, hadronization), and then supplemented by non-perturbative soft physics for the underlying event and multiple parton interactions.

5.4.1 Parton shower algorithm

It is beyond the scope of these lectures to provide a complete discussion of all the many ingredients that define a MC event generator. Instead, we focus here on a specific ingredient: the parton shower algorithm, which is rooted in perturbative QCD. The basic idea is that we want to generate a final state with **high partonic multiplicity**, to match what is found in a real collision. Instead of using exact perturbative calculations, we will exploit a crucial property in QCD which was first observed in our discussion of the $e^+ + e^- \to$ hadrons process:

- The probability for gluon radiation off a quark is enhanced when the gluon is either **soft** or **collinear**.
- For such a gluon, the probability of radiation is universal and **factorizable** from the rest of the process.
- We can then simulate multiple soft and collinear emissions, which represent the bulk of the cross-section, **independent of each other**, greatly simplifying the simulation of high-multiplicity final states.

Discussion #5.8

As described above, the parton shower process starts from the hard scattering matrix element, with the virtuality being $Q \gg \Lambda_{\rm qcd}$, and continues until the virtuality is of the order of $Q \sim \Lambda_{\rm qcd}$. Taking this into account, should we be worried with **soft and collinear divergences** present in the amplitudes of the process? Or instead these divergences get somehow regulated within the Monte Carlo simulation algorithm?

Let us now illustrate how the parton branching process is performed. A **splitting** of parton a into partons b and c, all of them massless, is depicted in Fig. 5.11. We will work under the assumption that the parton branching procedure is ordered in virtually

$$p_b^2 \, , p_c^2 \ll p_a^2 \equiv t \, , \tag{5.58}$$

which makes sense since the (time-like) branching process moves from large virtualities (result of the hard

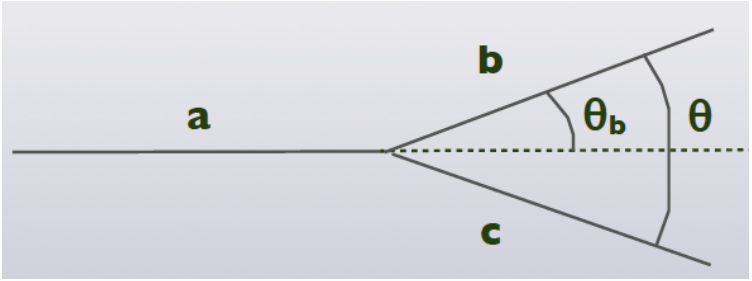


Figure 5.11. The collinear branching of a *mother* parton a into two *daughter* partons b and c.

scattering) to small virtualities (where non-perturbative QCD become relevant). Moreover, we will work in the small angle limit, where QCD matrix elements are enhanced due to infrared singularities, which as mentioned above is the same **collinear** limit that we have studied in the case of DIS and e^+e^- annihilation processes. In this limit the virtuality t reads

$$t = (p_b + p_c)^2 = 2\, E_b\, E_c\, (1 - \cos\theta) \simeq z(1-z)\, E_a\, \theta^2\,, \qquad z \equiv \frac{E_b}{E_a} = 1 - \frac{E_c}{E_a} \tag{5.59}$$

in terms of the energy fraction z and the branching angle θ. In this approximation, using also the conservation of transverse momentum in the splitting, we find the following useful relation:

$$\theta = \frac{\theta_b}{1-z} = \frac{\theta_c}{z}\,. \tag{5.60}$$

It is an interesting exercise to compute the splitting amplitudes for various possible combinations of quarks and gluons. In the case of the $g \to gg$ splitting, using the QCD Feynman rules, summarized in Fig. 2.3, we find that the vertex term can be written as

$$V_{ggg} = i g_s f^{def}\, [g^{\mu\nu}\,(p_a + p_c)^\rho + g^{\nu\rho}\,(p_c - p_b)^\mu - g^{\rho\mu}\,(p_a + p_b)^\nu] \tag{5.61}$$

Squaring the matrix element, we end up with a **factorized** expression for the matrix elements with $n+1$ particles $|\mathcal{M}_{n+1}|^2$, namely

$$|\mathcal{M}_{n+1}|^2 \sim \frac{4\,g^2}{t} C_A\, F(z; \epsilon_a, \epsilon_b, \epsilon_c) |\mathcal{M}_n|^2 \tag{5.62}$$

in terms of the matrix element with n particles $|\mathcal{M}_n|^2$ (computed before the splitting) and a universal factor F which depends only on the details of the splitting, but *not* of the hard scattering n-body matrix element. Summing over polarizations, we find that the structure of infrared divergences of the soft and collinear branchings are the same regardless of the hard scattering process: this is a consequence of the universality of soft and collinear singularities in QCD. We can write the $(n+1)$-particle squared matrix element $|\mathcal{M}_{n+1}|^2$ after the splitting as follows

$$\sum_{\text{pol}} |\mathcal{M}_{n+1}|^2 \sim \frac{4\,g^2}{t}\, C_A \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z)\right] |\mathcal{M}_n|^2 \tag{5.63}$$

We again observe this important property of QCD, namely that the cross-section for $n+1$ particles, when any pair has arisen from a soft or collinear splitting, can be factorized into the n particle cross-section times a universal factor, a splitting function, which is process independent.

From Eq. (5.63) we can observe the characteristic infrared divergences of the branching

- **Soft singularities** when $z \to 1$ or $z \to 0$: one of the partons becomes soft.

- **Collinear singularity** when $t \to 0$, see Eq. (5.59), when the two partons become collinear.

Finally, we can compute the cross-section for the $(n+1)$-particles final state in terms of the n-particle final state supplemented by a collinear branching, obtaining the following result,

$$d\sigma_{n+1} \,=\, d\sigma_n \, \frac{dt}{t} \, dz \, \frac{\alpha_s}{2\pi} \, \hat{P}_{ab}(z) \tag{5.64}$$

where $\hat{P}_{ab}(z)$ is the **unregularized splitting function**, a close sibling of the DIS splitting functions from Eq. (4.47), and is given, in particular case of a final-state $g \to gg$ splitting, by

$$\hat{P}_{gg}(z) = C_A \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right] , \tag{5.65}$$

and analogous expressions exist for other possible QCD branchings, such as $g \to q\bar{q}$ or $q \to qg$. Eq. (5.64) demonstrates how I can evaluate the probability of having a final state with multiplicity $n+1$ in terms of the probability of having a final state with multiplicity n, with the universal splitting functions dictating the relationship.

By including the contribution of more and more sequential branchings, we can compute the multiple emission of small-angle partons to all orders in the QCD perturbative expansions. This is what is done by the Monte Carlo parton showers such as `Pythia`, `Herwig` or `Sherpa`, hence we say that these showers have LL accuracy (meaning Leading Log). Therefore, parton showers are not only a method to simulate realistically the final state of high energy collisions, they also allow to improve the perturbative accuracy as compared to LO fixed order computations.

5.4.2 The parton shower MC branching equations

Let us now derive the Monte Carlo equations for a generic parton shower, starting from the hard scattering configuration. For simplicity, we will consider a *space-like shower*, corresponding the process schematically represented in Fig. 5.12: starting from a parton with low virtuality t_0 (since it is found in the proton, thus $t_0 \simeq \Lambda_{\rm QCD}$), it emits soft and collinear radiation under reaching the hard scattering virtuality $t = q^2$. The initial parton carries momentum fraction x_0, and the parton that enters the hard-scattering carries momentum fraction x_n.

The branching process depicted in Fig. 5.12 can be better represented by a path followed in the (t, x) plane, shown in Fig. 5.13. The way to compute the equations that govern the multiple soft and collinear branchings in a QCD parton shower is based on determining the number of paths that lead to a given region in the (t, x) plane and the number of paths that leave this same region and reach the hard scattering virtuality Q.

Using the schematic diagram in Fig. 5.13, the variation of the PDF due to all paths that enter into the (t, x) region is, using the result that we have derived in Eq. (5.64) for a soft or collinear branching, is the following:

$$\delta \, f_{\rm in}(x,t) \,=\, \frac{\delta \, t}{t} \, \int_x^1 \, dx' \, dz \, \frac{\alpha_s}{2\pi} \, \hat{P}_{qq}(z) \, f(x',t) \, \delta(x - x' \, z) \, , \tag{5.66}$$

where the PDF $f(x', t)$ is now inside the integral, while the contribution from the paths that leave this same region is given by

$$\delta \, f_{\rm out}(x,t) \,=\, \frac{\delta \, t}{t} \, f(x,t) \, \int_0^x \, dx' \, dz \, \frac{\alpha_s}{2\pi} \, \hat{P}_{qq}(z) \, \delta(x - x' \, z) \, , \tag{5.67}$$

where now the starting point of the PDF is fixed and thus is appears outside the integral. Therefore, putting them together, we obtain that due to the soft and collinear branchings, the PDF should satisfy an LL

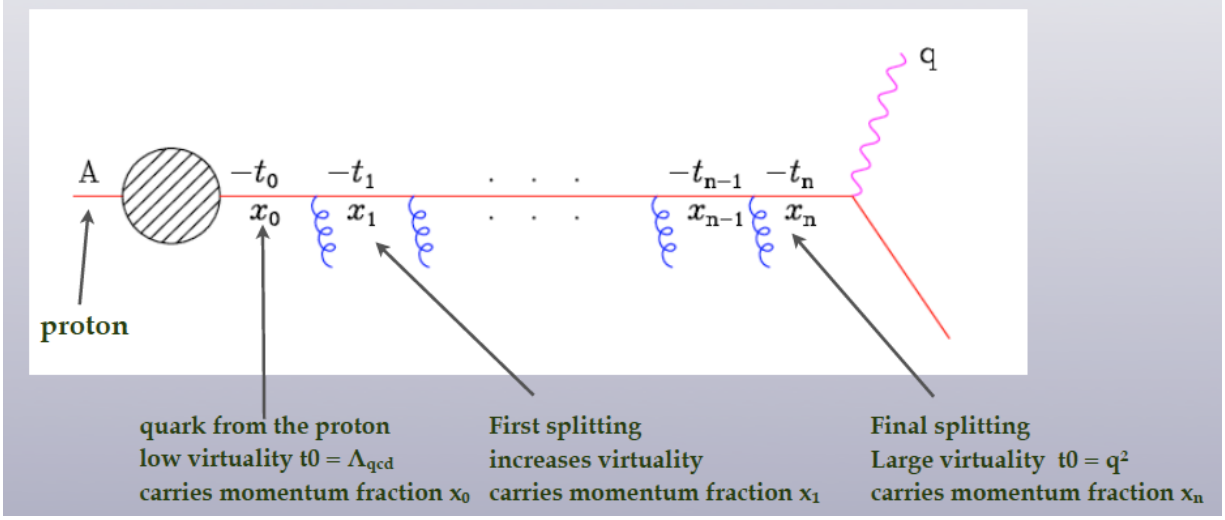


Figure 5.12. Schematic representation of a time-like parton shower branching, starting from a parton with low virtuality t_0, which emits soft and collinear radiation under reaching the hard scattering virtuality.

evolution equation:

$$t\frac{\partial}{\partial t}f(x,t) = \int_x^1 \frac{dz}{z}\frac{\alpha_s}{2\pi}P_{qq}(z)f\left(\frac{x}{z},t\right) \tag{5.68}$$

which is nothing but the same splitting functions that we encountered in DIS. Indeed, the parton branching description is an alternative way of deriving the QCD evolution equations.

Now we can introduce an important concept for Monte Carlo event generators, known as the **Sudakov form factor**. This quantity is defined as

$$\Delta(t) \equiv \exp\left(-\int_{t_0}^t \frac{dt'}{t'}\int dz\,\frac{\alpha_s}{2\pi}\hat{P}_{qq}(z)\right). \tag{5.69}$$

Using the Sudakov form factor, it is easy to derive the following modified evolution equations:

$$t\frac{\partial}{\partial t}\left(\frac{f(x,t)}{\Delta(t)}\right) = \frac{1}{\Delta(t)}\int\frac{dz}{z}\frac{\alpha_s}{2\pi}\hat{P}_{qq}(z)\,f\left(\frac{x}{z},t\right) \tag{5.70}$$

which can be easily integrated to give the following result:

$$f(x,t) = \Delta(t)\,f(x,t_0) + \int_{t_0}^t \frac{dt'}{t'}\frac{\Delta(t)}{\Delta(t')}\int\frac{dz}{z}\frac{\alpha_s}{2\pi}\hat{P}_{qq}(z)\,f\left(\frac{x}{z},t'\right) \tag{5.71}$$

Let's take a closer look to this equation

- The first term of the contribution to all paths where virtuality changes from t_0 to t without any branching
- The second term is the contribution from all paths that have a branching at t' and then evolve without branching up to a virtuality t

So the Sudakov form factor, Eq. (5.69) has a clean theoretical interpretation as the probability that, in a parton shower, a given parton evolves from t_0 to t without any branching (and then a branching takes place at t).

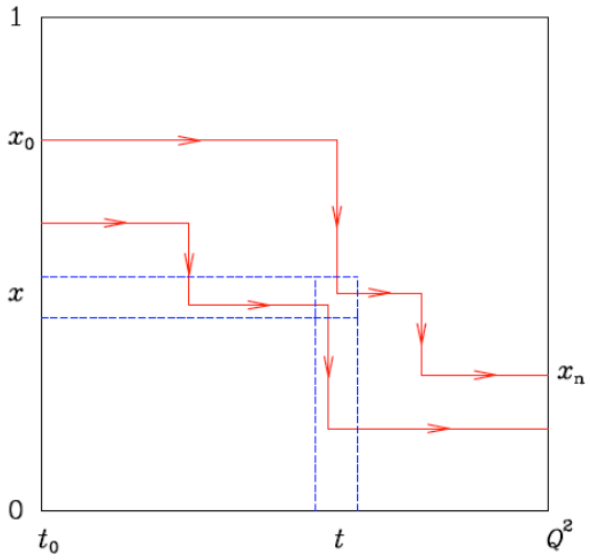


Figure 5.13. Path in the virtuality/momentum fraction plane for a initial state parton in a *space-like shower* from a low virtuality $t_0 \simeq \Lambda^2_{\rm QCD}$ to the hard scattering virtuality $t \simeq Q^2$. to high virtuality. See text for more details.

Eq. (5.69) is the basis of a Monte Carlo branching algorithm. The numerical implementation proceeds as follows

(a) At some point in the branching chain, we have a parton with virtuality t_1 and momentum fraction x_1,

(b) The MC algorithm generates, probabilistically, the values of t_2 using the equation

$$\frac{\Delta(t_2)}{\Delta(t_1)} = \mathcal{R} \tag{5.72}$$

where $\mathcal{R}$ is a flat random number between 0 and 1.

(c) Next the algorithm determines the value of x_2 of this splitting, using the condition

$$\int_\epsilon^{x_2/x_1} dz \, \frac{\alpha_s}{2\pi} \hat{P}_{qq}(z) = \mathcal{R}' \int_\epsilon^{1-\epsilon} dz \, \frac{\alpha_s}{2\pi} \hat{P}_{qq}(z) \tag{5.73}$$

where $\mathcal{R}'$ is another flat random number between 0 and 1, and the integral is performed only over **resolvable branchings**.

Using this procedure, starting from a relatively simple hard-scattering event, we can provide a realistic simulation of the final state particles using perturbative QCD. In Fig. 5.10 we show the schematic representation of the various ingredients that enter an hadronic collision, from the initial state (parton distributions) to the hard scattering cross-section to the final state (parton shower, hadronization), and then supplemented by non-perturbative soft physics for the underlying event and multiple parton interactions.

Once all partons in the shower have low virtualities, $t \sim \Lambda^2_{\rm QCD}$, perturbative QCD breaks down, and it is non-perturbative, long-distance dynamics that determine how these low-virtuality quarks and gluons are confined into the hadrons that are observed in the detector. The basic hypothesis is that of local quark-hadron duality: the flow of momenta and quantum numbers at hadron level follows that of the parton level. With this assumption, and taken into account the fundamental symmetries of QCD, various hadronization models have been proposed, and are implemented in different shower MC programs. For example, the cluster model exploits an important property of the pre-confinement of color in parton branching: color

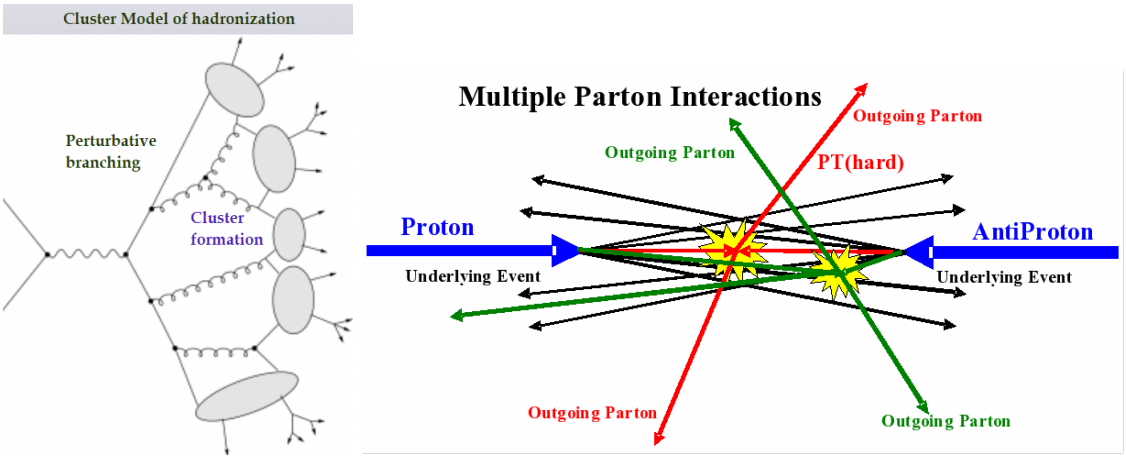


Figure 5.14. Left plot: schematic representation of the cluster model of hadronization, where a cluster formation step follows the perturbative parton branching process. Right plot: schematic representation of multiple parton interactions at an hadron collider.

singlet clusters of partons form after perturbative branching and then decay into the observed hadrons. This model is illustrated in Fig. 5.14, and the basic idea is to cluster the low-virtuality partons into color-singlet clusters that are then hadronized.

Finally, let us mention that real hadron-collider events need to deal not only with a single hard scattering, but actually in each event there can be multiple soft, semi-hard and hard collisions on top of the main one. If these collisions arise from the same basic hadron-hadron collision, they are known as **Multiple Parton Interactions**, and they are represented in Fig. 5.14. When collisions arise from different pp collisions within the same bunch-crossing, as happens when the collider luminosity is high, then they are denoted as *pile-up*. Taming the large pile-up present at the relatively large luminosities of the LHC is a major concern in experimental analysis.

5.5 Jet reconstruction and substructure

When discussing e^+e^- annihilation we found that exclusive observables exhibited infrared divergences, that could only be canceled in inclusive enough observables by means of a jet definition. At the LHC, jets are omnipresent, and jet reconstruction a crucial topic. In this lecture first we provide a general introduction to the topic of jet reconstruction, with emphasis on hadron colliders, and then we explore jet substructure, which is essential at the LHC specially in searches for massive new particles.

5.5.1 Jet reconstruction

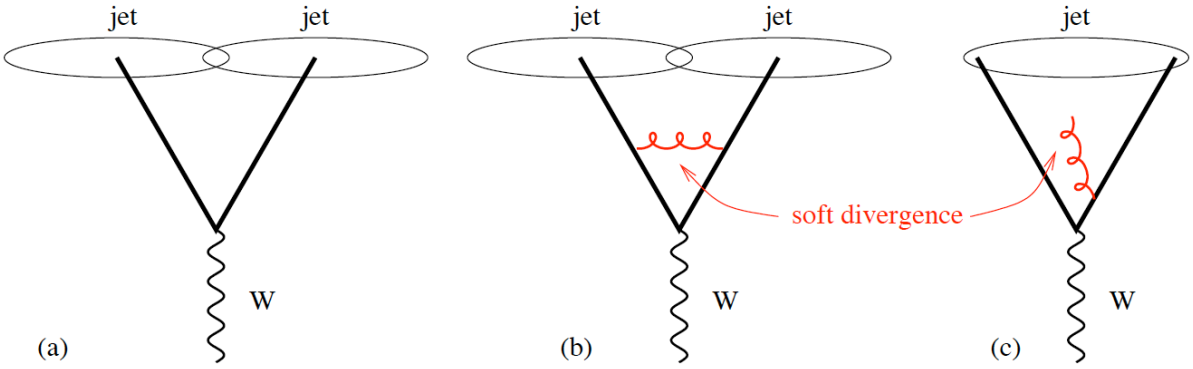
As we have seen in e^+e^-, jets are fundamentally ambiguous concept. A sensible jet definition needs to be introduced, that can be equally well applied to partons, hadrons and calorimeter cells. The conditions for a robust jet algorithm were defined in 1990 in the Snowmass accord:

Several important properties that should be met by a jet definition are [3]:

1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields finite cross sections at any order of perturbation theory;
5. Yields a cross section that is relatively insensitive to hadronisation.

A crucial property of a jet definition is that it should be *infrared and collinear safe*, that is, provide sensible results to all orders in the QCD perturbative expansion. As we saw in the case of the Sterman-Weinberg jets, infrared safety can be formulated by requiring that a cross-section, computed in pQCD, should be unchanged if any of the particles undergoes a soft or a collinear splitting.

An example of a IRC unsafe definition would be the following:



since we see the emission or an arbitrarily soft parton merges the two jets that would otherwise be separated, spoiling the cancellation of real and virtual soft divergences and producing an infinite result. The most popular jet reconstruction algorithm at the LHC is known as the *anti-k_T algorithm*. The is to cluster pairs of partons sequentially following some measure of their *distance*. The algorithm works as follows

(a) Make a list of all final-state particles in your collision (can be partons, hadrons, calorimeter cells)

(b) Now compute the following distance between a pair of particles

$$d_{ij} = \min\left(p_{T,i}^{-2}, p_{T,i}^{-2}\right) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \,. \tag{5.74}$$

where R is the jet radius, the fundamental parameter of the algorithm. We also compute the distance of particle i with the beam

$$d_{iB} = p_{T,i}^2 \,. \tag{5.75}$$

(c) Find the minimum distance. If it is a d_{ij}, recombine the two particles by adding their four-momentum. Else, declare particle i to be a jet and remove it from the list of particles.

(d) Continue the algorithm until the list of particles is empty

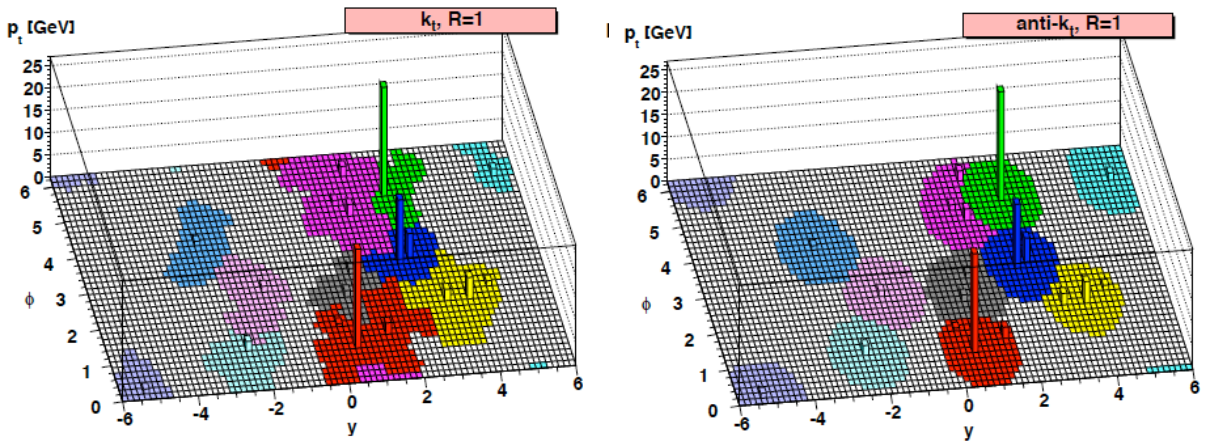


Figure 5.15. The catchment area of two jet algorithms, with $R = 1$: the k_T algorithm and the anti-k_T algorithm. It is clear that the second produces much more regular jets than the first of the algorithms.

An important property of the anti-k_T algorithm is that its *catchment area* (the area in which each each picks up soft particles) is relatively regular, and this is a very beneficial property for jet energy calibration as well as for the subtraction of underlying event and pileup. In Fig. 5.15 we show the catchment area of two jet algorithms, with $R = 1$: the k_T algorithm and the anti-k_T algorithm. It is clear that the second produces much more regular jets than the first of the algorithms. The k_T algorithm is defined in exactly the same way as the anti-k_T algorithm but with a modified distance

$$d_{ij} = \min\left(p_{T,i}^2, p_{T,i}^2\right) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \,. \tag{5.76}$$

Another widely used sequential recombination algorithm is know as the Cambridge/Aachen algorithm, where the distance is based purely on angular variables

$$d_{ij} = \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \,, \tag{5.77}$$

and which is specially useful in the context of jet substructure, as we discuss now.

5.5.2 Jet substructure

A closely related topic that has seen an impressive development recently is that of jet substructure. The motivation to look closer to the internal structure of jets is provided by the following observation. Consider the decays of a resonance with a mass at the electroweak scale. Typically due to the limited phase space it will be produced with a small boost, and thus, if decays to QCD partons, the resulting jets will be well separated. However, at the LHC a new kinematical regime becomes available: that where the EW-scale resonance is produced with a very large boost, $p_T \gg m$. In these conditions, the subsequent hadronic decays of the resonance end up being collimated into a single jet, becoming thus indistinguishable from the overwhelming QCD background - unless jet substructure tools are used.

The situation is illustrated in Fig. 5.16. A heavy resonance, such as a W boson or a top quark, produced with a large enough boost, $p_T \gg m$, when decaying hadronically the final state becomes indistinguishable from than of QCD jets. This seems a major problem, since it seems one would be missing many events of important electroweak processes because of the dominance of the QCD jet backgrounds. Given that many BSM scenarios involve the production of boosted heavy resonances, a major limitation to the LHC program

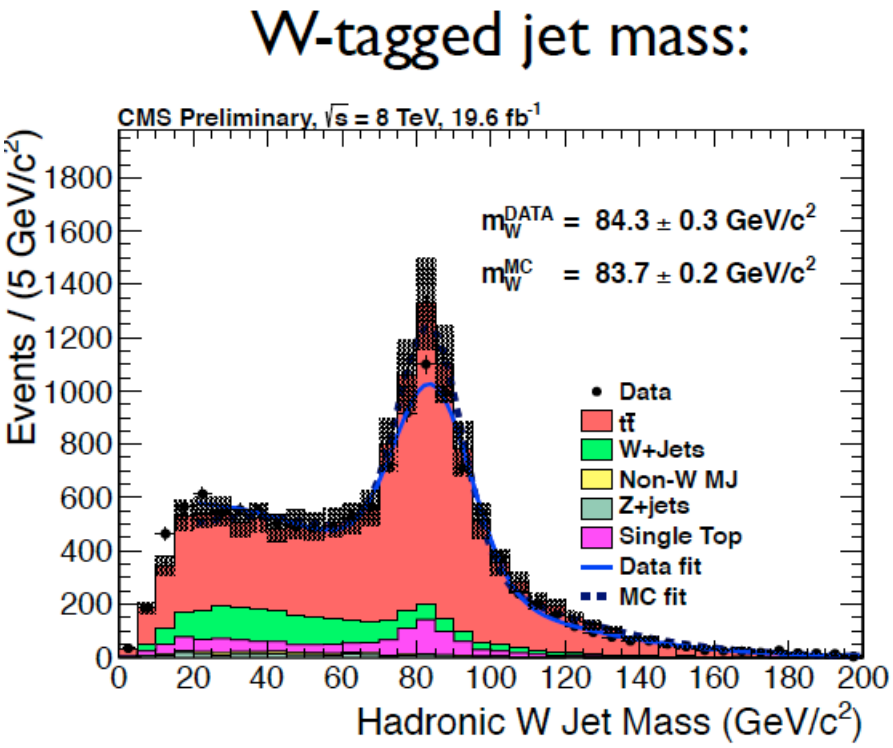


Figure 5.16. If a heavy resonance is produced with a large enough boost, $p_T \gg m$, then when decaying hadronically the final state becomes indistinguishable from than of QCD jets - unless one looks at the jet substructure.

might be feared. Fortunately, we can use our understanding of QCD radiation to distinguish background from interesting signals even in the boosted regime.

To understand which is the kinematic regime relevant for jet substructure let's consider one of the most paradigmatic examples, namely the decay of a boosted Higgs boson into a $b\bar{b}$ pair. The processes is illustrated schematically in Fig. 5.17. For simplicity, we assume that the Higgs has been produced centrally, and take the azimuthal angle to be $\phi = 0$.

Using the parametrisation which is more suitable for hadronic collisions, we can write the four-momentum of the Higgs boson as follows:

$$p_H = \left(\sqrt{m_H^2 + p_{T,H}^2}\,, p_{T,H}\,, 0\,, 0\right) \tag{5.78}$$

and that of the decay bottom quarks

$$p_b = \left(z\sqrt{m_H^2 + p_{T,H}^2}\,, p_{x,b}\,, 0\,, p_{z,b}\right) \tag{5.79}$$

$$p_{\bar{b}} = \left((1-z)\sqrt{m_H^2 + p_{T,H}^2}\,, p_{x,\bar{b}}\,, 0\,, -p_{z,b}\right) \tag{5.80}$$

where z is the fraction of the Higgs boson original energy that is being carried by the b quarks.

Now using four-momentum conservation and neglecting the small masses of the bottom quarks, we find that the kinematics of the process is uniquely defined in terms of three variables:

- The Higgs boson mass, m_H,
- The Higgs transverse momentum, $p_{T,H}$,
- and the fraction of energy carried by the bottom quark z (which determines the asymmetry of the

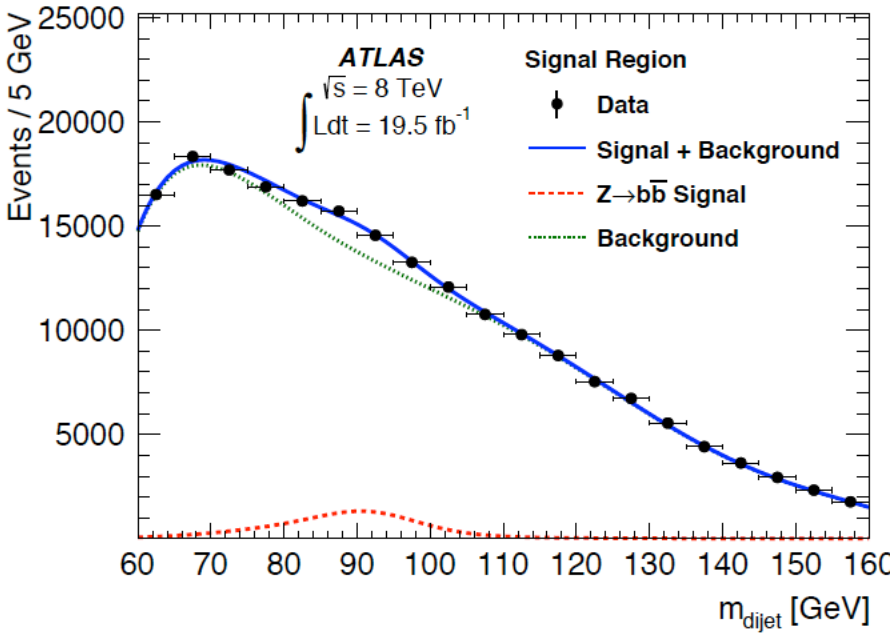


Figure 5.17. Scheme for the central production of a Higgs boson in a hadronic collision, followed by the decay into a $b\bar{b}$ pairs.

decay)

It is an easy computation to show that

$$p_{x,b} = \frac{p_{T,H}}{2} + (2z-1)\frac{p_{T,H}^2+m_H^2}{2\,p_{T,H}} \tag{5.81}$$

$$p_{x,\bar{b}} = \frac{p_{T,H}}{2} - (2z-1)\frac{p_{T,H}^2+m_H^2}{2\,p_{T,H}} \tag{5.82}$$

$$p_{z,b} = \left[(p_{T,H}^2+m_H^2)\left(z^2-z+\frac{1}{2}\right) - \frac{p_{T,H}^2}{4} - (2z-1)^2\frac{(p_{T,H}^2+m_H^2)^2}{4\,p_{T,H}^2}\right]^{1/2} \tag{5.83}$$

The above formulae are valid in all generality. Now let us consider the boosted regime, $p_{T,H} \gg m_H$, and then we obtain a substantial simplification to end up with

$$p_{x,b} = z\,p_{T,H} \tag{5.84}$$

$$p_{x,\bar{b}} = (1-z)\,p_{T,H} \tag{5.85}$$

$$p_{z,b} = \sqrt{z\,(1-z)}\,m_H \tag{5.86}$$

Therefore, in the boosted regime, the angular separation between the bottom and anti-bottom quarks will be given by

$$R_{b\bar{b}} = \frac{p_{z,b}}{p_{x,b}} - \frac{p_{z,\bar{b}}}{p_{x,\bar{b}}} = \frac{1}{\sqrt{z\,(1-z)}}\frac{m_H}{p_{T,H}} \tag{5.87}$$

Eq. (5.87) allows to determine when the boosted regime kicks in at the level of final state topology. If we have a jet definition with jet radius R, them, if the transverse momentum of the Higgs boson is such that $R \geq R_{b\bar{b}}$, the two b quarks will end up collimated into a single jet, and there traditional Higgs reconstruction, based on two separate b-tagged jets, becomes impossible. For $R = 0.5$ and $z = 1/2$, we find that in the case

of $H \to b\bar{b}$ the boosted regime is relevant when $p_{T,H} \geq 500$ GeV. This of course is a reason of concern, since these high $p_{T,H}$ events are very important for Higgs characterization studies.

To be able to find a way out, a number of jet substructure techniques have been developed in order to efficiently distinguish jets with non-trivial substructure from QCD jets. The basic ideas are common

- QCD jets arise from soft and collinear splittings, while in the decay of heavy resonances all prongs share a similar amount of energy
- QCD jets do not show structure in the jet mass distribution, while for heavy resonances the jet mass is peaked around the resonance mass.
- Removing soft radiation in a jet should leave signal events unchanged, while decreasing the contamination from QCD events

A plethora of jet substructure techniques have been developed in the recent years. Here we will discuss one of the most important substructure taggers, known as the BDRS mass-drop tagger. This tagger is based on the Cambridge/Aachen algorithm. In the Cambridge/Aachen algorithm, first of all compute all possible angular distances between all objects that enter your jet definition

$$\Delta R_{ij}^2 \, = \, (y_i - y_j)^2 \, + \, (\phi_i - \phi_j)^2 \tag{5.88}$$

First of all we require a substantial mass drop, so that the subjet masses are much smaller that the original jet mass

$$m_{j1} \, \leq \, \mu \, m_j \tag{5.89}$$

and then that the splitting is not too asymmetric

$$y_{\rm mdt} \, \equiv \, \frac{\min(p_{T,j1}, p_{T,j2})}{m_j^2} \Delta R_{j1,j2j}^2 \, \geq \, y_{\rm cut} \tag{5.90}$$

$$m_j^2 \, = \, m_{j1}^2 \, + \, m_{j2}^2 \, + \, 2\, z(1-z)\, E_j^2 \, \theta_{j1,j2}^2 \tag{5.91}$$

$$m_{j1}^2 \, \simeq \, m_j^2 \left(1 - \frac{2\, E_j^2}{m_j^2} \, \theta_{j1,j2}^2 \right) \, \leq \, \mu \, m_j^2 \tag{5.92}$$

Now, in the collinear limit, using the formulae that we have derived previously, it is easy to see that

$$y_{\rm mdt} \, \equiv \, \frac{\min(p_{T,j1}, p_{T,j2})}{m_j^2} \Delta R_{j1,j2j}^2 \, \simeq \, \frac{\min\left(z_{j1}^2, z_{j2}^2\right)}{z_{j1}\,, z_{j2}} \, = \frac{\min\left(z_{j1}, z_{j2}\right)}{\max\left(z_{j1}\,, z_{j2}\right)} \, \geq \, y_{\rm cut} \tag{5.93}$$

In QCD, we typically have

$$z_{j1} \, \gg \, z_{j2} \quad \text{or} \quad z_{j1} \, \ll \, z_{j2} \tag{5.94}$$

which in the case of the decays of some heavy resonance, typically into two equal mass particles

$$z_{j1} \, \sim \, z_{j2} \tag{5.95}$$

Many other jet tagging algorithms have been proposed in the literature.