

# MC event generator introduction

Terascale Monte Carlo School  
November 08, 2021  
Stefan Prestel (Lund University)



The goal of these lectures is to

give you basic background on different aspects and algorithms in high-energy physics event generators

the lectures are split into two parts:

- General overview, basic sampling algorithms, phase space, hard scattering, veto algorithm
- Showers, MPI, hadronization

I aim for a broad, but not too detailed overview. Overlap with the other lectures is expected :)

An observation in particle physics is

$$\langle O \rangle = \int d\phi_n \frac{d\sigma(A, B \rightarrow n \text{ particles})}{d\phi_n} O(\phi_n)$$

phase space      differential cross section      value of observable

**phase space**: sample of all quantum numbers (momentum, flavor...) of particles in scattering final state

**differential cross section**  $\approx$  transition probability to scattering final state

Compare to expectation value in statistics:

$$\langle g \rangle = \int dx f(x) g(x)$$

random variables      probability distribution      value of function

$\Rightarrow$  Calculate "theory predictions" for  $O$  with statistical methods.

- Dedicated calculations : Evaluate analytic expressions on paper...or very likely a computer. Safe & fast, but only viable for “simple” problems
- Monte Carlo generators : Approximate analytic expressions numerically, by statistical sampling on a computer. Use Monte-Carlo methods to handle complex scattering final states and/or observations.

Monte-Carlo algorithms are simple enough to have wide applicability, e.g. in integration

$$\int_{x_-}^{x_+} dx f(x) = (x_+ - x_-) \langle f \rangle \approx \frac{(x_+ - x_-)}{N} \sum_{i=1}^N f(x_i)$$

The approximation errors is  $\propto \frac{1}{\sqrt{N}}$ , independent of number of integrations  
( $dx \rightarrow dx_1 \cdots dx_n$ )

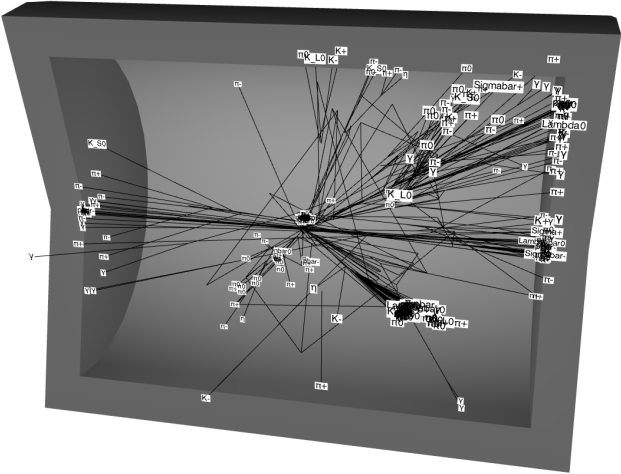
Ideally suited for our types of integrals

$$\langle O \rangle = \int d\Phi_n \frac{d\sigma_n}{d\Phi_n} O(\Phi_n) \propto \frac{1}{N} \sum_{i=1}^N \frac{d\sigma_n}{d\Phi_n}(\Phi_n^{(i)}) O(\Phi_n^{(i)})$$

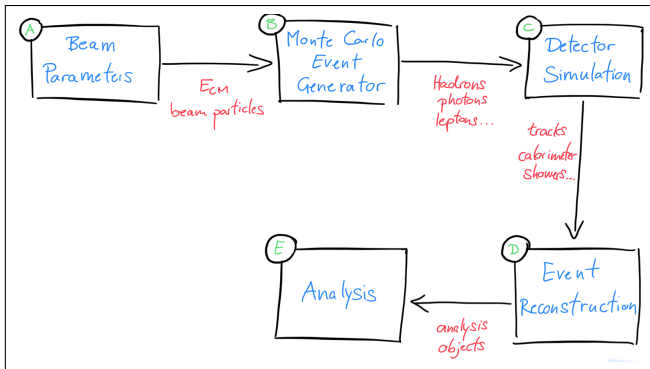
May even store the **events**  $\Phi_n^{(i)}$  with **event weight**  $\frac{d\sigma_n}{d\Phi_n}(\Phi_n^{(i)})$  and evaluate  $O(\Phi_n^{(i)})$  later!

NB: Les Houches Event Files are effectively that.

You can think of an event is several ways...



The sampling (=event generation) of complicated phase space points  $\Phi_n^{(i)}$ , and the calculation of  $\frac{d\sigma_n}{d\Phi_n}(\Phi_n^{(i)})$  can (with some theory, and some hand-waving) be factorized into smaller problems:



The Monte-Carlo generator landscape is rich! Just to name a few:

Neutrino physics:

Genie, GiBUU, NuWro, NEUT...

Cosmic rays:

EPOS, QGSJET and SIBYLL

Heavy ions:

HIJING, AMPT, JEWEL...

LHC physics:

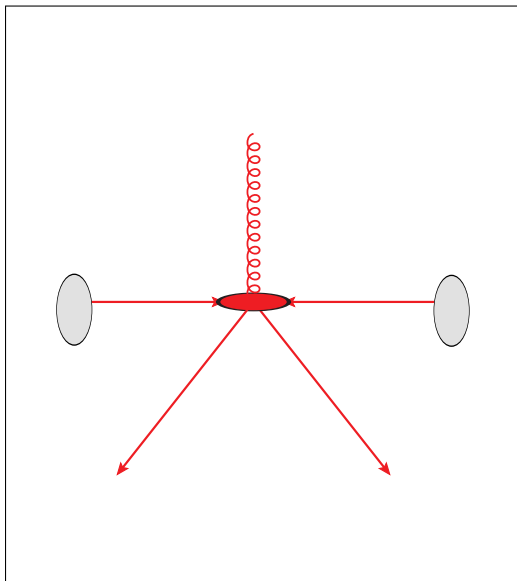
Herwig, Pythia, Sherpa  
Madgraph, Whizard, Alpgen...

KaTie, Cascade

All of them amazing tools to learn about phenomenology!

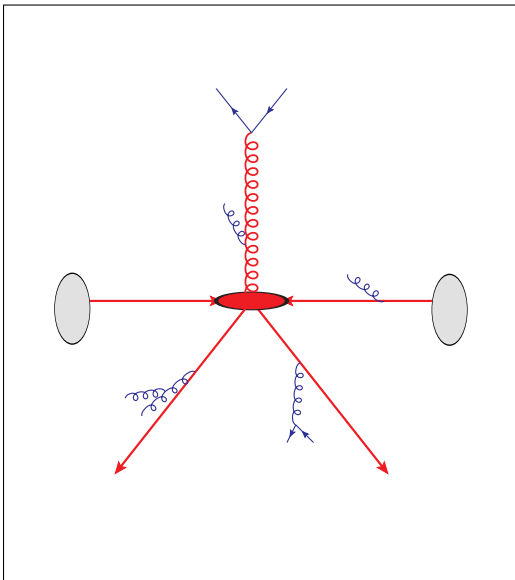


A high-energy scattering breaks the beams apart



A high-energy scattering breaks the beams apart

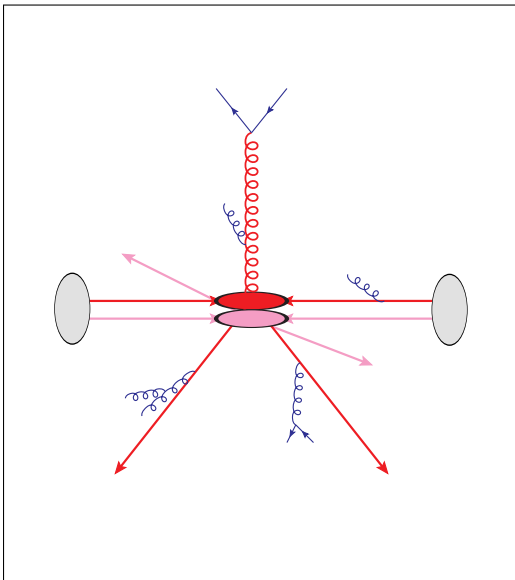
...which initiates a cascade of radiation in the vacuum.



A high-energy scattering breaks the beams apart

...which initiates a cascade of radiation in the vacuum.

Secondary interactions might occur at the same time

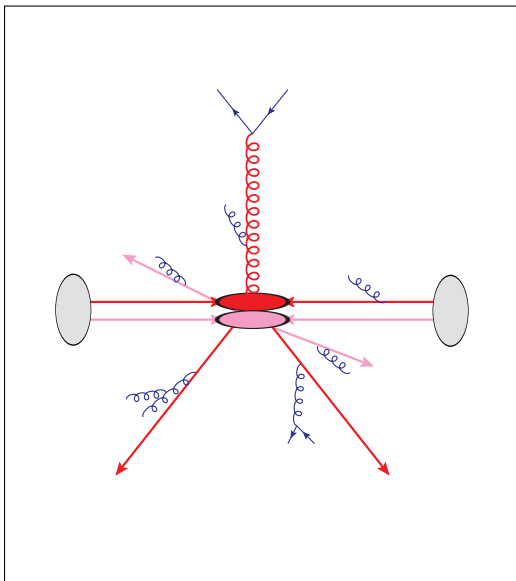


A high-energy scattering breaks the beams apart

...which initiates a cascade of radiation in the vacuum.

Secondary interactions might occur at the same time

...and initiate further radiation "showers".



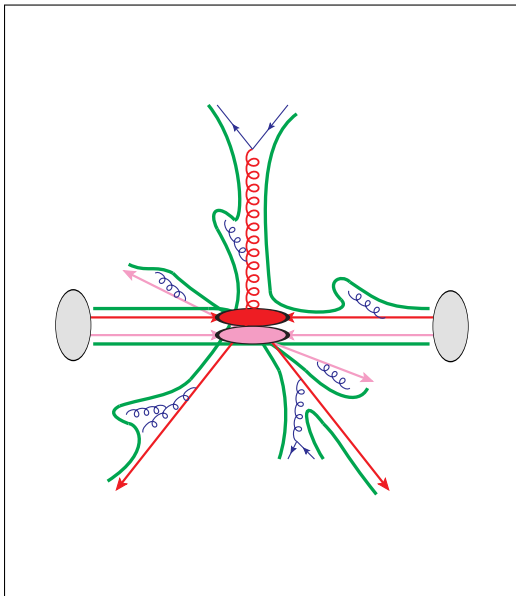
A high-energy scattering breaks the beams apart

...which initiates a cascade of radiation in the vacuum.

Secondary interactions might occur at the same time

...and initiate further radiation "showers".

Confining potentials form, once the  $\langle E \rangle$  per particle is small



A high-energy scattering breaks the beams apart

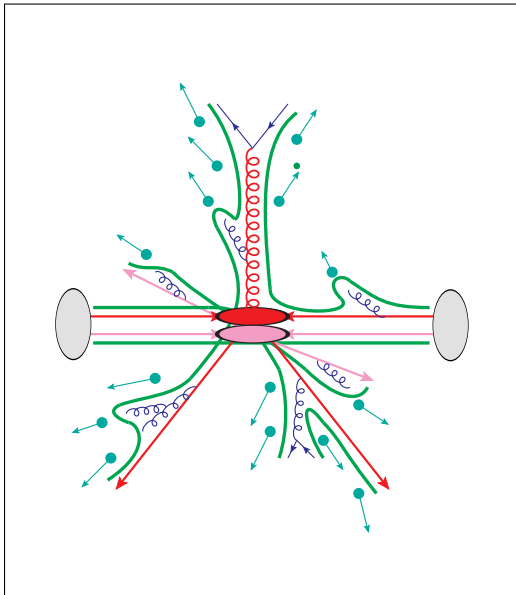
...which initiates a cascade of radiation in the vacuum.

Secondary interactions might occur at the same time

...and initiate further radiation "showers".

Confining potentials form, once the  $\langle E \rangle$  per particle is small

...leading to the nucleation of excited or unstable hadrons



A high-energy scattering breaks the beams apart

...which initiates a cascade of radiation in the vacuum.

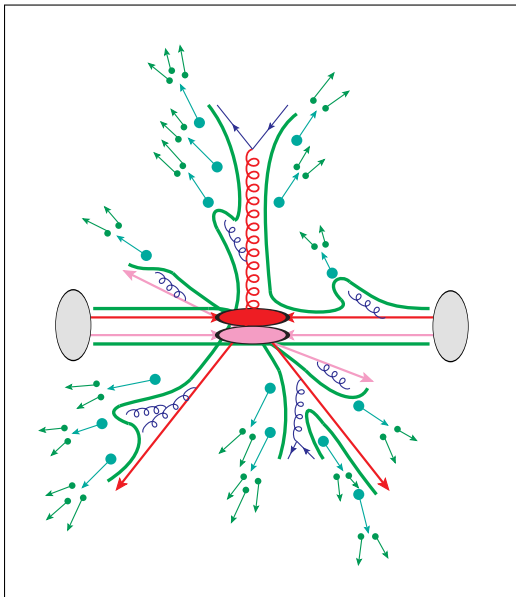
Secondary interactions might occur at the same time

...and initiate further radiation "showers".

Confining potentials form, once the  $\langle E \rangle$  per particle is small

...leading to the nucleation of excited or unstable hadrons

...which decay into stable states



A high-energy scattering breaks the beams apart

...which initiates a cascade of radiation in the vacuum.

Secondary interactions might occur at the same time

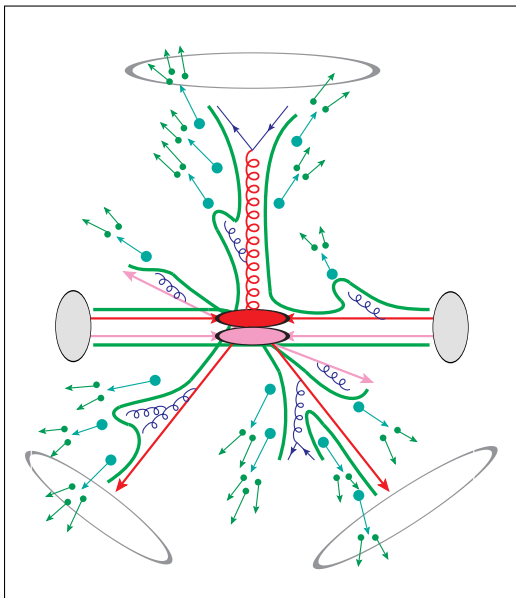
...and initiate further radiation "showers".

Confining potentials form, once the  $\langle E \rangle$  per particle is small

...leading to the nucleation of excited or unstable hadrons

...which decay into stable states.

[outside MCEG: interactions with the detector material occur, analysis objects are reconstructed]





From a technical viewpoint, this chain of phenomena looks like

$$\begin{aligned} dP(\text{beams} \rightarrow \text{final state}) \\ &= dP(\text{beams} \rightarrow A, B) \\ &\otimes dP(A, B \rightarrow \text{few partons}) \\ &\otimes dP(\text{few parton} \rightarrow \text{many partons}) \\ &\otimes dP(\text{many partons} \rightarrow \text{hadrons}) \\ &\otimes dP(\text{hadrons} \rightarrow \text{stable particles}) \end{aligned}$$

Very high integration dimension. Traditionally, only Monte-Carlo viable  
→ Need to learn about numerical methods

An overview of some basic numerical techniques gives a feeling about how to tackle event generation.

Nowadays, deep nets can be used to simulate special cases.

Assume we want to **pick a random variable according to a distribution** (e.g. a phase-space point)

The cumulative distribution is

$$C(y) = \int_{-\infty}^y dx p(x) \quad \text{with} \quad \int_{-\infty}^{\infty} dx p(x) = 1$$

which allows using  $R \in [0, 1]$  and

$$C(y) = R \quad \Rightarrow \quad y = C^{-1}(R)$$

This is called inversion sampling.

Often, we're not so lucky that a uniquely **invertible primitive function**  $C^{-1}$  exists ...but we can often still use this method as part of a more flexible algorithm.

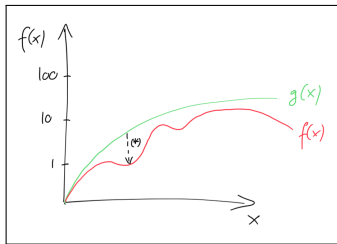
We can circumvent the issue with rejection sampling (a.k.a. hit-or-miss)

Basic idea: Use a simple distribution to pick  $x$  from, adjust rate once  $x$  is generated.

- Assume a simple distribution  $g(x) > f(x)$ , i.e.

$$f(x) = g(x) \underbrace{\frac{f(x)}{g(x)}}_{< 1}$$

- Use inversion sampling to draw  $x$  from  $g(x)$ .
- Draw  $R \in [0, 1]$ . Reject  $x$  if  $\frac{f(x)}{g(x)} < R$



$\Rightarrow$  Accepted  $x$  now distributed according to  $f(x)$ . This algorithm is excessively used in Monte Carlo generators.

Let's get back to physics for a bit :)  
The measurement of an observable is

$$\langle O \rangle = \int d\Phi_n \frac{d\sigma(A, B \rightarrow n \text{ particles})}{d\Phi_n} O(\Phi_n)$$

phase space      differential cross section      value of observable

...so we have to worry about

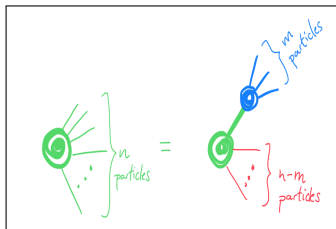
- sampling phase space points  $\Phi_n$
- calculating the differential cross section  $\frac{d\sigma_n}{d\Phi_n}$
- evaluating the observable

When sampling phase space,  
 avoid large event weight fluctuations  
 avoid excessive rejection rate

⇒ Phase space generation separates enthusiasts from experts.

$$d\Phi_n = \left[ \prod_{i=1}^n \frac{d\vec{p}_i}{(2\pi)^3 2E_i} \right] \delta(p_A + p_B - \sum_1^n p_i)$$

This  $(3n - 4)$  dimensional integration can be sampled in factorized steps:



$$d\Phi_n = d\Phi_{n-m+1} \frac{ds_{1m}}{2\pi} d\Phi_m$$

...we can continue until only simple integrations ( $d\Phi_2, d\Phi_3$ ) remain, and then find a **clever parameterization** for those → Use knowledge about  $d\sigma$ : Importance sampling!

$$\begin{aligned}
 \left| \text{Diagram} \right|^2 &= \left| \text{Diagram}_{P_1, P_3} \right|^2 + \left| \text{Diagram}_{P_3, P_2} \right|^2 + \text{Interference} \\
 f &= g_1 + g_2 \\
 &\quad \text{enhanced for } (P_1 + P_3)^2 \rightarrow 0 \quad \text{enhanced for } (P_2 + P_3)^2 \rightarrow 0
 \end{aligned}$$

Differential cross sections have a rich structure. In that case, importance sampling can be **combined** with the discrete transformation method into multichannel sampling:

- Use  $f(x) \leq g_1(x) + g_2(x)$
- Choose index  $i \in \{1, 2\}$  [using  $P_i = \int dx g_i(x)$ ]
- Draw  $x$  from  $g_i(x)$ . Overall,  $x$  is now distributed according to  $g_1 + g_2$
- Draw  $R \in [0, 1]$ , and accept if  $(i, x)$  pair if  $\frac{f(x)}{g_1(x) + g_2(x)} < R$ . Else reject & restart.

NB: also heavily used in parton showers.

**Exercise:** Draw  $x$  from the distribution  $f(x) = \frac{1}{\sqrt{x(1-x)}}$  using two integration channels.

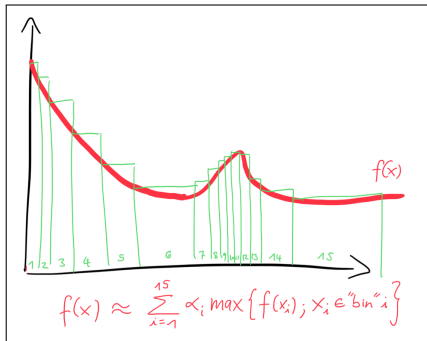
All of these methods require (analytical) knowledge of the differential cross section – which is often hard to come by.

Another way of “generating variables in integration regions where they matter most” is stratified sampling:

- Multichannel with  $g_i \propto \max\{f\}$  in small integration region (=bin).
- Put more bins where variance of  $f(x)$  is large.

This is the construction principle of VEGAS.

NB: Need to evaluate the function very often to learn good “integration grids”.



Phase-space integrators in MCs are a mix of all of these methods, and recently also more modern machine learning techniques.

Once we have a phase-space point, it's time to evaluate the differential cross section

$$\frac{d\sigma}{d\Phi_n} = \int \underbrace{dx_A dx_B f_A(x_A, M_F^2) f_B(x_B, M_F^2)}_{\substack{\text{distribution of interacting} \\ \text{particles in beams} \\ \Rightarrow \text{Parametrized} \\ \text{measurement}}} \frac{1}{F} |\mathcal{M}|^2$$

↑  
Transition  
probability

The calculation of the **transition probability**  $|\mathcal{M}|^2$  relies on perturbative methods:

**Pen & paper:** Calculate Feynman diagrams,  
use completeness relations to square,  
sum over external quantum numbers (helicity, color...)

**Real life:** Assemble helicity amplitudes for fixed color  
add & square  $\Rightarrow$  less complicated intermediate expressions, better scaling



Tree-level calculations on their own are questionable: **Beware** of how to count coupling powers (and particle number).

Infrared (IR) singularities abound in tree-level diagrams ...because the “particle number” operator is **ill-defined** in perturbative QFT!

Singularities cancel between *different multiplicities* when introducing virtual corrections.

Notes:

- The result are *inclusive* cross sections.
- Measurements that ensure singularity cancellation are called *IR safe*.

Example :

$\mu$   $\mu^*$   $O(g^2)$

... divergent due to loop integral

$\mu$   $\mu^*$   $O(g^2)$

... divergent due to extra phase space integral

KLN theorem : Divergences cancel!

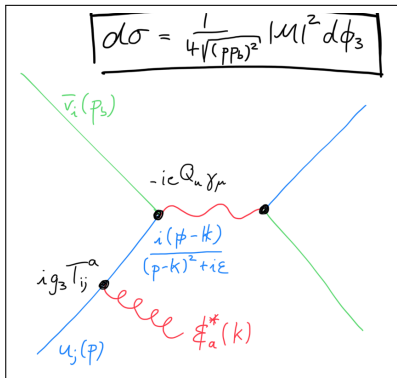
Infrared singularities in multi-parton amplitudes have a profound consequence: Nature will dress partons with many more partons to take advantage of the enhancement!

For small  $p_{\perp \text{gluon}}$  and  $E_{p-k} \approx z E_p$ , the internal quark is almost on-shell, and

$$\frac{i(\not{p} - \not{k})}{(p-k)^2 + i\epsilon} \approx \frac{u(p_a)\bar{u}(p_a)}{p_a^2}$$

$$d\Phi_3 \approx d\Phi_2 \frac{d\phi dz dp_{\perp}}{4(2\pi)^2(1-z)}$$

$$\frac{1}{4\sqrt{(pp_b)^2}} \approx z \frac{1}{4\sqrt{(p_a p_b)^2}}$$



All components of the x-section factorize, and we're left with

$$d\sigma_3 \approx d\sigma_2 \int d\phi dz dp_{\perp} P(\phi, z, p_{\perp})$$

where the universal splitting function  $P$  contains the singularities due to gluon emission.

Once the divergences have been factorized, we may attempt to calculate an observable to next-to-leading order accuracy

$$\begin{aligned} \langle O \rangle_{\text{NLO}} &= \int d\Phi_n \left\{ \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} + \frac{d\sigma_n^{\text{Virt}}}{d\Phi_n} + \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes \int d\Phi_1 S \right\} O(\Phi_n) \\ &+ \int d\Phi_{n+1} \left\{ \frac{d\sigma_{n+1}^{\text{Tree}}}{d\Phi_{n+1}} O(\Phi_{n+1}) - \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes SO(\Phi_n) \right\} \end{aligned}$$

where  $\frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes S$  captures the singularities of real-emission and – **by the KLN theorem** – virtual corrections alike.

This allows numerical predictions for IR-safe observables, i.e. when  $O_{n+1} \rightarrow O_n$  when the additional particle becomes unresolvable.

However, it **does not allow** the generation of “NLO events”.

Remember the KLN theorem: Infrared singularities arising in real-emission diagrams cancel against alike divergences in virtual corrections.<sup>1</sup>

For the (most) enhanced parts, we can devise a radical interpretation of KLN:

$$\int dk \text{ (virtual) } \begin{array}{c} \text{diagram} \\ \otimes \theta(\text{cuts}) \end{array} = - \int dk \text{ (real) } \begin{array}{c} \text{diagram} \\ \otimes \theta(\text{cuts}) \end{array}$$

The diagram shows a mathematical equation enclosed in a rectangular box. On the left side, the word "virtual" is written in blue above a Feynman diagram. The diagram consists of two vertices connected by a curved line labeled with momentum  $k$ . Each vertex has two external lines. Below the diagram is the expression  $\otimes \theta(\text{cuts})$  in red. To the right of this is an equals sign followed by a minus sign. On the right side, the word "real" is written in blue above another Feynman diagram. This diagram is similar to the virtual one but includes an additional external line representing a real emission. Below this diagram is also the expression  $\otimes \theta(\text{cuts})$  in red.

“The rate for # particles remaining the same is (negative) the rate for the # particles increasing at any scale  $t$  – even in the presence of cuts/regularization”.

This is the first building block of a parton shower.

<sup>1</sup> This is a popularized account; there are subtleties. Kinoshita’s paper highly recommended.

The behavior of partons is similar to that of radioactive elements.

The # particles  $n$  can only change  $n \rightarrow n + 1$  (due to decay or splitting) at scale  $t$  if it has not already changed at  $t' > t$ .

The probability to not change in a finite interval  $\Delta t$  is

$$1 - \Delta t P(t)$$

where  $P$  is the splitting kernel containing the enhanced parts of the real correction. This is simply statement about unitarity: The rate of no change and the rate of all possible changes add to unity.

The probability not to change in any very small sub-interval  $\Delta t/n$  is

$$\left(1 - \frac{\Delta t}{n} P(t)\right)^n \xrightarrow{n \rightarrow \infty} \exp\left(-\int_0^{\Delta t} dt P(t)\right)$$

This exponential suppression of not splitting is called the Sudakov factor.

[no splitting]  $\leftrightarrow$  [fixed # particles]. Thus, the Sudakov introduces virtual corrections.

Combined, the decay/splitting probability at scale  $t$  is

$$\mathcal{P}(t) = P(t) \exp\left(-\int_0^t d\bar{t} P(\bar{t})\right) = P(t)\Delta(t)$$

Retains a memory: the “next” decay may only happen at scale  $t$  if it had not happened before. **Conservation of total probability means that the process develops a “memory”.**

Note that this means that the no-decay probability follows the differential equation

$$\underbrace{-\frac{d\Delta(t)}{dt}}_{\text{change of \#particles by decay}} = P(t)\Delta(t) \quad \leftrightarrow \quad -\frac{d \ln \Delta(t)}{dt} = P(t)$$

It is possible to rewrite the DGLAP equation in this form:

$$\int_x^1 \frac{dz}{z} \frac{\alpha_s}{2\pi} \frac{1}{t} [P(z)]_+ f\left(\frac{x}{z}, t\right) \quad \longleftrightarrow \quad \frac{d \ln(\Pi(x, t_{\text{max}}, t))}{dt} = \int_x^{1-\varepsilon} \frac{dz}{z} \frac{\alpha_s}{2\pi} \frac{1}{t} P(z) \frac{f\left(\frac{x}{z}, t\right)}{f(x, t)}$$

We can use differential equation to define an inversion sampling algorithm that correctly includes the “memory”:

$$-\frac{d \ln \Delta(t)}{dt} = P(t) \quad , \quad \Delta(t) = \exp \left( - \int_0^t P(t) \right) = \exp (-F(t) + F(0))$$

Note that  $\Delta(t)$  is the cumulative function of  $\frac{d\Delta}{dt}$ , i.e. of the probability density that defines the distribution of  $t$  values. Thus, draw  $R \in [0, 1]$  and

$$R = \Delta(t) = \exp (-F(t) + F(0)) \quad \Rightarrow \quad t = F^{-1}(F(0) - \ln R)$$

...and we've produced a sample of decay scales (with memory). This is the basic algorithm used in parton showers.

**In this way, parton showers can solve evolution equations.** The result incorporates exponential Sudakov factors, i.e. is an all-order “resummed” prediction.

However, for most cases of interest,  $F^{-1}$  does not exist – rejection sampling to the rescue. However, it's important to **retain the memory**.

This is achieved by the Sudakov veto algorithm:

- o Assume a simple distribution  $g(t) > f(t)$ , i.e.  $f(t) = g(t) \frac{f(t)}{g(t)}$
- 1 Set  $t_0 = 0$
- 2 Use inversion sampling to draw  $t$  from  $g(t)$  (using  $t_0$  as lower bound).
- 3 Draw  $R \in [0, 1]$ . Reject  $t$  if  $\frac{f(t)}{g(t)} < R$   
Wrong: Restart at 1 ← this would erase the memory!  
Correct: Set  $t_0 \rightarrow t$ , restart at 2.

In this way, parton showers can solve complicated evolution equations.

In nature, many different “decay channels” ( $g \rightarrow q\bar{q}, g \rightarrow gg$ ) may compete  $\Rightarrow$  Add algorithm to treat competition (often winner-takes-all)

NB: Typically, the algorithm is rearranged to move from large  $t$ -values ( $\mathcal{O}(\mu_f)$ ) to small  $t$ -values ( $\mathcal{O}(1\text{GeV})$ ).



With this, we're finally able to construct a parton shower, since

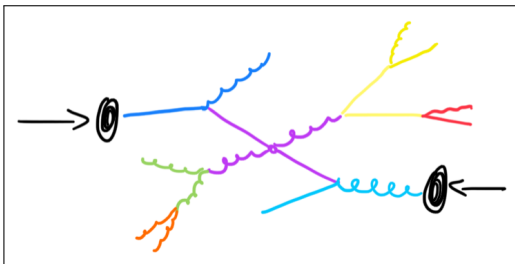
- Within the simplest approximation, the splitting functions are universal, and fully factorized from the “hard” cross section
- Within the simplest approximation, decays are independent (apart from being ordered in a decreasing sequence of scales)

⇒ The splitting process can be iterated, with the result after  $n$  splittings forming the “hard” scattering for the  $(n + 1)$ th emission.

The effect of the shower  $\mathcal{F}$  on an observable  $O$  is, symbolically,

$$\begin{aligned} \mathcal{F}_n(O, \Phi_n, t_{\max}, t_{\min}) &= \Delta_n(t_{\max}, t_{\min})O(\Phi_n) \\ &+ \int_{t_{\min}}^{t_{\max}} d\Phi_1 \Delta_n(t_{\max}, t)P(\phi, z, t)\mathcal{F}_{n+1}(O, \Phi_{n+1}, t, t_{\min}) \end{aligned}$$

Through  $\Delta$ , the shower is an “all-order” calculation, and each term in the formula is individually finite.



The parton shower will develop from high propagator virtuality and large angles to small virtuality and angle.

Several choices will influence the sequence: **how are the emissions ordered?** **how is the phase space for emissions mapped?** **how are quantum interferences approximated?**

The most prominent features of the event will be determined by the highest-momentum transfer transition

⇒ best to calculate that exactly.

⇒ don't let the shower to produce that transition.

Maybe the most natural evolution variable is **propagator virtuality**.

The transition matrix element for gluon radiation from a  $q\bar{q}$  pair is

$$\begin{aligned}
 |\mathcal{M}_{X+g}|^2 &\sim |\mathcal{M}_X|^2 \alpha_s \frac{2(p_q p_{\bar{q}})}{(p_q p_g)(p_g p_{\bar{q}})} \\
 &\sim |\mathcal{M}_X|^2 \alpha_s \frac{1}{E_g^2} \frac{(1 - \cos \Theta_{q\bar{q}})}{(1 - \cos \Theta_{qg})(1 - \cos \Theta_{\bar{q}g})} \\
 &= |\mathcal{M}_X|^2 \alpha_s \frac{1}{E_g^2} \frac{1}{2} \overbrace{\left[ \frac{(1 - \cos \Theta_{q\bar{q}})}{(1 - \cos \Theta_{qg})(1 - \cos \Theta_{\bar{q}g})} + \frac{1}{1 + \cos \Theta_{qg}} - \frac{1}{1 - \cos \Theta_{\bar{q}g}} \right]}^{W_q} \\
 &+ \quad q \leftrightarrow \bar{q}
 \end{aligned}$$

Radiation from the quark is determined by  $W_q$ , and

$$\int \frac{d\phi_{qg}}{2\pi} W_q = \begin{cases} 1/(1 - \cos \Theta_{qg}) & \text{if } \Theta_{qg} < \Theta_{q\bar{q}} \\ 0 & \text{else} \end{cases}$$

The angle of emissions (from  $q$ ) is decreasing.

The effect of **angular ordering** can be incorporated

- Ordering emissions in angle
- Ordering emissions in transverse momentum\*
- Ordering emissions in virtuality *and implementing the correct transition  $ME^\dagger$*

The choice of ordering variable has physical consequences:

- angular ordering restricts available phase space
- transverse-momentum ordering may not easily recover (subleading) color correlations

There is no “magic bullet”. The choice of ordering variable  $\oplus$  splitting kernel really is **an uncertainty of the shower construction**

\* due to properties of Lorentz boosts; only guaranteed for one emission

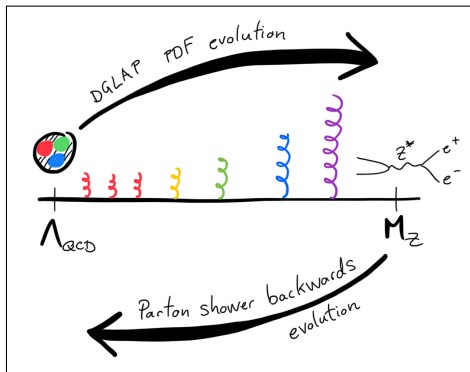
† only works for one emission, not suitable in general

Radiation from incoming particles *change* that particle's momentum, i.e. **change the probability to extract this parton from a hadron beam**

Initial-state radiation has to consider parton distribution functions. PDFs **change/evolve** with hardness scale.

The highest-momentum transfer transition should not be generated by the generated by the shower. Thus, it is best to **work from the highest scale backwards**.

⇒ **Initial-state showers require a re-write the PDF evolution equations.**



You'll hear a lot more about **forward** and **backward** evolution in the next days.

The presence of (analytic) PDF evolution constraints is one of the main reasons why *parton showers beyond leading order are hard*.

For example, there are currently no showers that can handle NLO PDFs consistently<sup>†</sup>.

It is however possible to correct event generators **order-by-order** or **emission-by-emission**.

<sup>†</sup> this would require handling of negative splitting “probabilities” and double-emission...which might require new factorization theorems as foundation.

Compare a next-to-leading order calculation and an expanded version of the shower:

$$\begin{aligned}
 \langle O \rangle_{\text{NLO}} &= \int d\Phi_n \left\{ \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} + \frac{d\sigma_n^{\text{Virt}}}{d\Phi_n} + \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes \int d\Phi_1 S \right\} O(\Phi_n) \\
 &+ \int d\Phi_{n+1} \left\{ \frac{d\sigma_{n+1}^{\text{Tree}}}{d\Phi_{n+1}} O(\Phi_{n+1}) - \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes SO(\Phi_n) \right\} \\
 \langle O \rangle_{\text{PS}} &= \int d\Phi_n \left\{ \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} - \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \int_{t_{\min}}^{t_{\max}} d\Phi_1 P(\phi, z, t) + \mathcal{O}(\alpha^2) \right\} O(\Phi_n) \\
 &+ \int_{t_{\min}}^{t_{\max}} \int d\Phi_n d\Phi_1 \left\{ \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} P(\phi, z, t) + \mathcal{O}(\alpha^2) \right\} O(\Phi_{n+1})
 \end{aligned}$$

As expected, the calculations overlap (the shower gives an approximation of NLO).

Suggestion: Subtract the PS result from the NLO, and use the result as starting point of the shower, instead of  $\frac{d\sigma_n^{\text{Tree}}}{d\Phi_n}$

The big advantage of this suggestion is that we can (finally!) generate **NLO events** – just add a couple for zeros:

$$\begin{aligned}
 \langle O \rangle_{\text{NLO}} = & \int d\Phi_n \left\{ \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} + \frac{d\sigma_n^{\text{Virt}}}{d\Phi_n} + \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes \int d\Phi_1 S \right. \\
 & \left. - \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \otimes \int d\Phi_1 S + \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \int_{t_{\min}}^{t_{\max}} d\Phi_1 P(\phi, z, t) \right\} O(\Phi_n) \\
 + & \int d\Phi_{n+1} \left\{ \frac{d\sigma_{n+1}^{\text{Tree}}}{d\Phi_{n+1}} - \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} P(\phi, z, t) \Theta(t_{\min}, t_{\max}) \right\} O(\Phi_{n+1}) \\
 + & \int d\Phi_n \frac{d\sigma_n^{\text{Tree}}}{d\Phi_n} \int_{t_{\min}}^{t_{\max}} d\Phi_1 P(\phi, z, t) \left\{ O(\Phi_{n+1}) - O(\Phi_n) \right\}
 \end{aligned}$$

Both  $\{\dots\}$  are separately finite.  $\{\dots\}$  is just the 1st-order expansion of the shower – which we would produce by showering the three first lines.

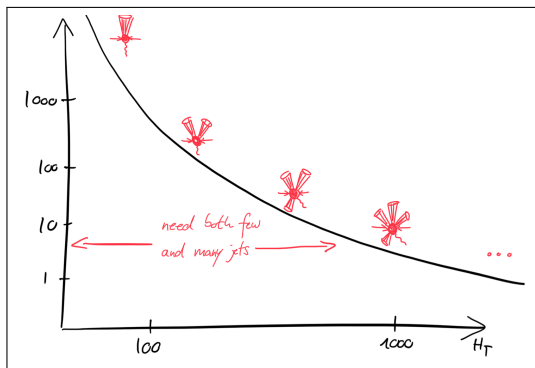
Removing  $\{\dots\}$  allows to generate events. Showering the result produces a consistent NLO matched calculation, in the MC@NLO approach.



NLO matched calculations will describe one additional jet with tree-level accuracy.

Analyses of experimental data often depend on multi-jet final states, e.g. to expose Beyond-the-SM signals.

In this case, NLO (or NNLO or N3LO) matching is often not sufficient.



Instead, consistently “stack” simpler (tree-level or NLO) calculations on top of each other, with the help of the shower. This defines a merging scheme.

The task for a tree-level merging scheme is to describe events for

[simple final state  $X$ ] +  $\{0, 1, \dots, N\}$  well-separated jets

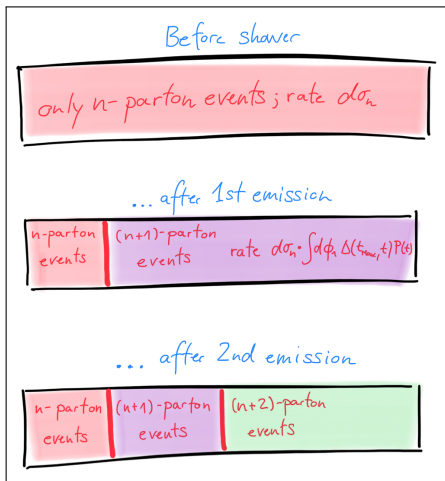
through a combined calculation, with tree-level accurate  $X + \{0, 1, \dots, N\}$  parton rates, and the jets' structure determined by the parton shower.

Simply adding several showered tree-level calculations is inconsistent, since the results overlap.

Take inspiration from PS to avoid overlap:

- Showers produce (all-order) real emission corrections
- The lower-multiplicity (inclusive) cross section is preserved by removing the emission rate from the rate of lower-multiplicity events.

An idealized merging method could handle overlap in exactly the same way.



The chain of reasoning is

$$\int d\Phi_n O(\Phi_n) \frac{d\sigma_n}{d\Phi_n} + \int d\Phi_{n+1} O(\Phi_{n+1}) \frac{d\sigma_{n+1}}{d\Phi_{n+1}} + \dots$$

make  $(n + 1)$  PS-like  $\rightarrow$

$$\int d\Phi_n O(\Phi_n) \frac{d\sigma_n}{d\Phi_n} + \int d\Phi_{n+1} O(\Phi_{n+1}) \frac{d\sigma_{n+1}}{d\Phi_{n+1}} \Delta_n(t_n, t_{n+1}) + \dots$$

remove real from Born  $\rightarrow$

$$\int d\Phi_n O(\Phi_n) \frac{d\sigma_n}{d\Phi_n} - \int d\Phi_{n+1} O(\Phi_n) \frac{d\sigma_{n+1}}{d\Phi_{n+1}} \Delta_n(t_n, t_{n+1}) + \int d\Phi_{n+1} O(\Phi_{n+1}) \frac{d\sigma_{n+1}}{d\Phi_{n+1}} \Delta_n(t_n, t_{n+1}) + \dots$$

make more PS-like  
 $\approx$

$$\int d\Phi_n O(\Phi_n) \frac{d\sigma_n}{d\Phi_n} \Delta_n(t_n, t_{\min}) + \int d\Phi_{n+1} O(\Phi_{n+1}) \frac{d\sigma_{n+1}}{d\Phi_{n+1}} \Delta_n(t_n, t_{n+1})$$

effective description  
 $\approx$

$$\int d\Phi_n O(\Phi_n) \frac{d\sigma_n}{d\Phi_n} [\text{veto events with more than } n \text{ hard jets}] + \int d\Phi_{n+1} O(\Phi_{n+1}) \frac{d\sigma_{n+1}}{d\Phi_{n+1}} [\text{veto events jets harder than in ME}] + \dots$$

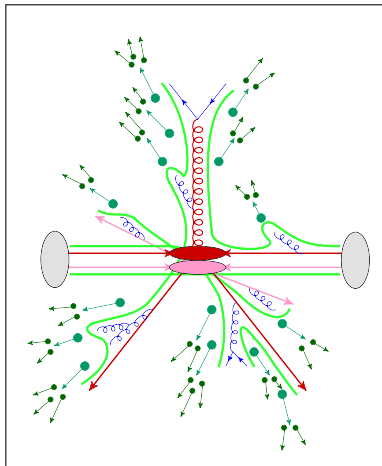
Several tree-level and NLO merging prescriptions have been implemented, with various approximations of “preserving the inclusive cross section”.

Let's take a step back, and look at the bigger picture.

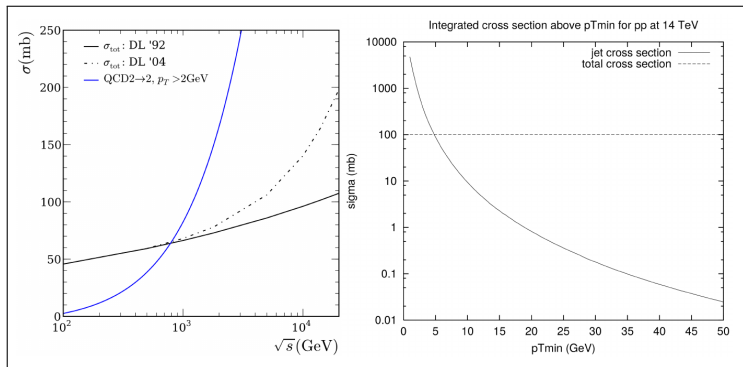
At hadron colliders, the initial state is complex.

There is no reason to expect only one parton-parton interaction to occur.

Does the inclusion of multiple interactions change the inclusive single-interaction cross section?



The naive inclusive cross section for parton-parton scattering is often divergent already at leading order.



This simply hints at a too literal interpretation of the concept of “inclusive cross section”.

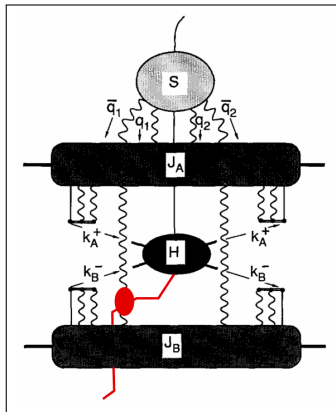
The crux lies in the definition of the parton distribution functions: These give the inclusive probability to find a parton at  $x$  with all other interactions above  $x \approx \frac{p_{\perp \min}}{E_{CM}}$  integrated out.

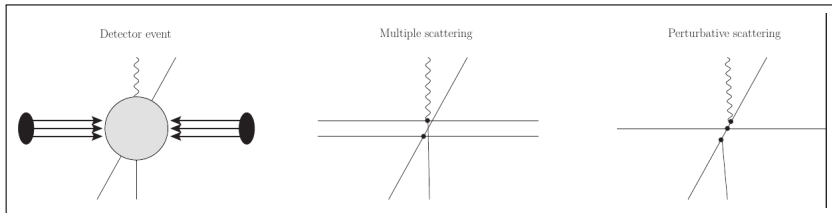
Detailed enough measurements will probe the integrand, i.e. be sensitive to multiple interactions.

In this case, we should interpret the cross section as

$$\sigma_{\text{inclusive}}(p_{\perp \min}, E_{CM}) = \langle n(p_{\perp \min}) \rangle \cdot \sigma_{\text{inelastic}}(p_{\perp \min}, E_{CM})$$

$$\sigma_{\text{inelastic}}(p_{\perp \min}, E_{CM}) < \sigma_{\text{total}}(E_{CM})$$





Take a four-jet event as an example:

- jets might not be separated and emerge from showering
- jets might be well-separated and emerge from one scattering
- jets might be well-separated and emerge from two scatterings

It is important to understand the measurement in order to understand the cocktail of phenomena.



**Argument:** Want inclusive x-section to be calculable in perturbation theory + PDFs. Multiple interactions should not change this. Simply overlaying scatterings will not work.

**Realization:** Multiple interactions are not additive – just as tree-level calculations are not!

**Solution:** The rate for not having a second interaction is correlated with the rate for having a second interaction.

Note the similarity to loops  $\leftrightarrow$  reals and shower emission rate  $\leftrightarrow$  Sudakov factor

Unitarity (= conservation of probability) suggests a phenomenological model:

$$\langle O \rangle = \int \mathcal{Y} O(\phi_1) - \int \mathcal{Y} O(\phi_1) \int_{t_{\min}}^{t_{\max}} \text{---} + \int \mathcal{Y} \int_{t_{\min}}^{t_{\max}} \text{---} O(\phi_1 \oplus \phi_2)$$

subtract  $\leftarrow$   $\rightarrow$  what you add

shower-like  $\rightarrow$

$$\int \mathcal{Y} \exp\left(-\int_{t_{\min}}^{t_{\max}} \text{---}\right) O(\phi_1) + \int \mathcal{Y} \int_{t_{\min}}^{t_{\max}} \text{---} \exp\left(-\int_t^{t_{\max}} \text{---}\right) O(\phi_1 \oplus \phi_2)$$

even more shower-like  $\swarrow$

$$\mathcal{F}_{\text{PS+M}}(O, \phi_1 \oplus \phi_2, t, t_{\min})$$

In fact, this is basically the same algorithm as for parton showering.

So as always, the proof is in the pudding

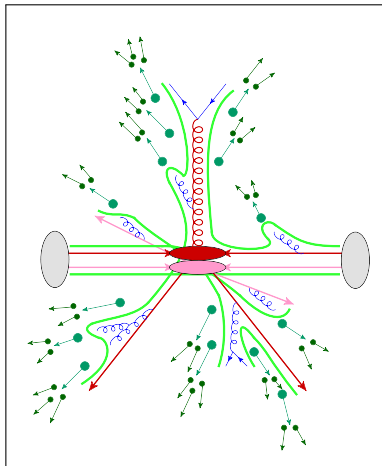
- no reason to expect primary and secondary partons to be in the “same place” in the proton  
Multiple interactions introduce impact parameter dependence
- some inelastic scattering cross sections (evaluated at fixed order) still require regularization for small momentum transfer
- the correlation and competition between multiple interactions and showers is non-trivial

We started from an overview of event generation at microscopic detail.

- $dP(\text{beams} \rightarrow \text{final state})$
- $= dP(\text{beams} \rightarrow A, B)$
- $\otimes dP(A, B \rightarrow \text{few partons})$
- $\otimes dP(\text{few parton} \rightarrow \text{many partons})$
- $\otimes dP(\text{many partons} \rightarrow \text{hadrons})$
- $\otimes dP(\text{hadrons} \rightarrow \text{stable particles})$

The **last steps** are typically responsible for a vast increase in particle multiplicity.

Phenomenological models & data parameterization are employed here.



Nobody has solved strong-coupling QFTs yet. Until then, we require a model to translate set of partons to sets of hadrons.

So how do partons coalesce?

**Individual partons**  
→ hadrons

...as e.g. introduced by  
Feynman & Field

What about flavor and  
momentum conservation?

Not ideal, but still partially  
used ( $\sim$  fragmentation  
functions)

**All partons**  
→ all hadrons

In conflict with per-  
turbative QCD (&  
non-universal)

Difficult to imagine  
“jetty” behavior.

Still useful for extremely  
high-multiplicity  $\oplus$  low  
 $\langle E \rangle$  events

**Subset of partons**  
→ subset of hadrons

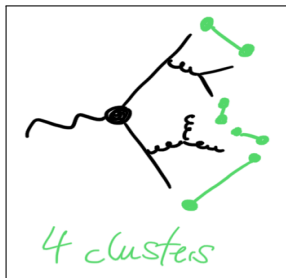
Middle ground between  
the extremes

Basis of the most success-  
ful high-energy physics  
models – the **string** and  
**cluster** model.

Main approach in Event  
Generators.

Partons “close to” each other hadronize coherently.

There are two main schools of thought of what “close to” means:



### Cluster hadronization

- create clusters from color-connected partons (gluons branch to two quarks)
- invoking **color preconfinement**

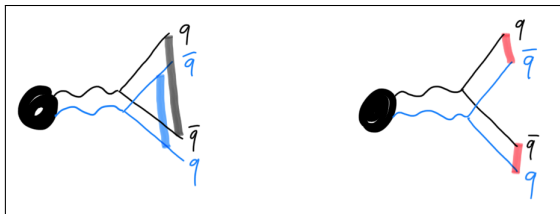
### String hadronization

- create strings from color string, with gluons “stretching the string” locally
- invoking **non-perturbative insights**

Note already here: real-life models borrow traits and phenomena from both – depending e.g. on available phase space for hadrons.

The notion of closeness determines which partons hadronize collectively.

In busy systems – like LHC collisions – definitions of closeness are typically less obvious



Previously independent systems might undergo color reconnection, e.g. to neutralize flavor more locally.

Color reconnection is not a completely random process: Minimizing some measure of energy ( $\sim \sum_{i,j \in \text{partons}} \ln(p_i p_j)$ ) is likely to occur.

It is an unspoken assumption of CR models that the total cross section is unaffected by any rearrangement.

Similarly, the transition **partons**  $\rightarrow$  **hadrons** does not change the total cross section, i.e. colored partons coalesce into hadrons with unit probability

Having discussed the sets of partons that collectively hadronize, we may now discuss the **string** (PYTHIA) and **cluster** (HERWIG, SHERPA) models.

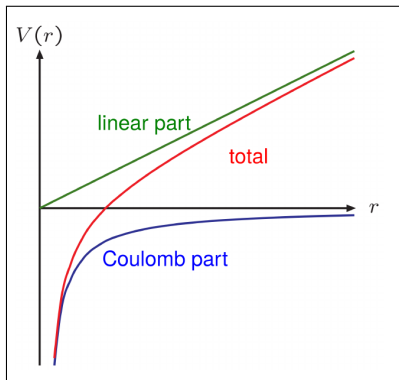


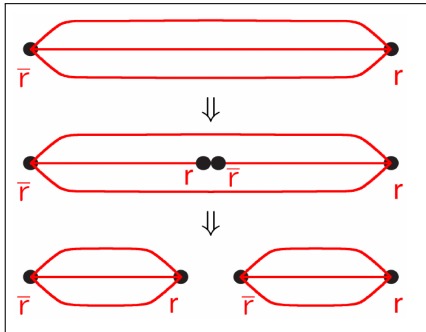
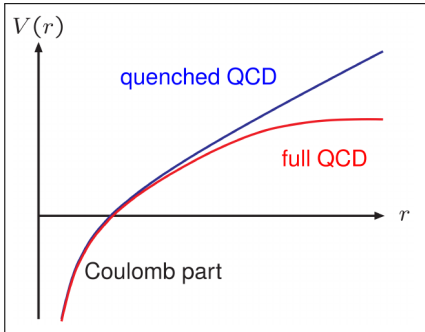
Although non-perturbative QCD is hard, some results are known e.g. from lattice QCD.

The potential between two quarks is linear, since the force per unit length is constant.

The force is confining, and similar to the force on a stretched string.

This is the basis of the **string model**.





In reality, the force between quarks will drop eventually: It is energetically favorable for the string to break.

Mesons are  $\approx$  oscillating strings – so-called yo-yo modes.

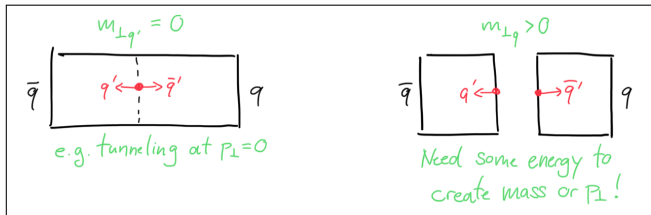
High-energy strings break through pair creation.

Strings break through  $f\bar{f}$  creation through a tunneling mechanism (Heisenberg & Euler, Schwinger – yes, that old).

QCD strings break through  $q\bar{q}$  creation with tunneling probability

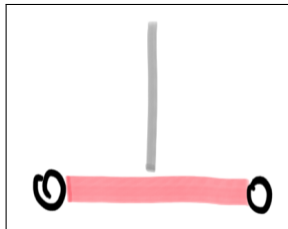
$$\mathcal{P} \propto \exp\left(-\frac{\pi m_{\perp q}^2}{\kappa}\right) \quad \kappa = \text{string tension}$$

Tunneling of heavy quarks suppressed by  $m_{\perp q}^2$  dependence.  $c\bar{c}$  almost negligible.



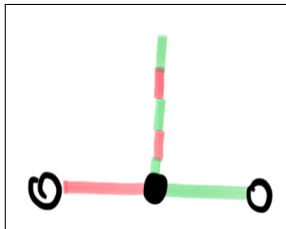
High transverse momentum suppressed. Breaking yields  $\approx$  back-to-back particle production in string CM frame.

QCD contains both quarks and gluons, i.e. realistic model should consider gluons as well.



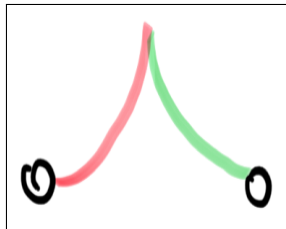
Gluon does not change color field.

Very unlikely



Gluon induces new type of string, attached by junction.

Adds new, unknown parameters.

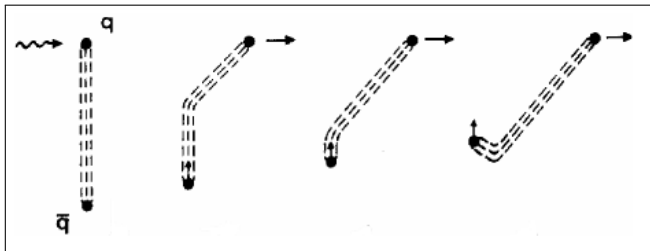


Gluon is a "kink" on the string.

Kinks are present on massless relativistic strings.

No additional parameters needed.

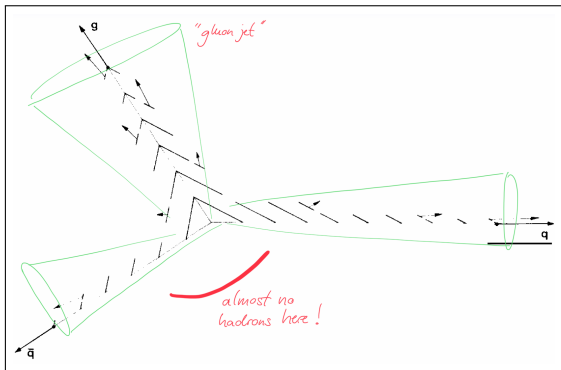
A “kink” is a large, instantaneous momentum transfer at the initial time. It stretches the string in some direction.



The kink is connected to two string segments. Thus, it **loses energy twice as fast as the endpoints**, in accordance with QCD, where  $C_A/C_F \xrightarrow{N_C \rightarrow \infty} 2$

Causality dictates that the string + kink system fragment like any other string.

The interpretation of gluons as kinks has an important consequence: the string effect



There are almost no hadrons in the region opposite the jet formed by the gluon kink.

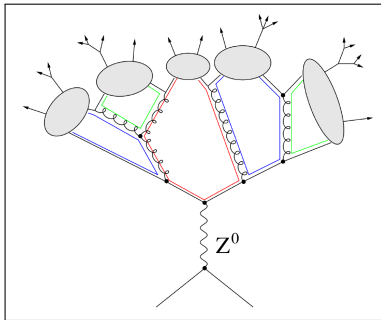
The gluon kink and the quark endpoints act coherently to deplete that region.

Coherence effects are already found in perturbative QCD: Gluon production at comparable angles is suppressed by destructive interference.

Thus, color-singlet parton pairs end up “close” in phase space. This is called preconfinement. Preconfinement mimics the string effect at perturbative level.

This is the basis of the cluster model:

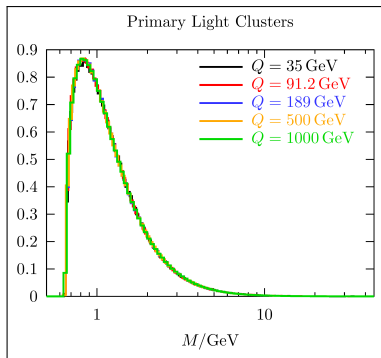
- use perturbative calculation that enforces coherence
- convert gluons to  $q\bar{q}$  pairs with heuristic model
- collect  $q\bar{q}$  pairs into color-singlet clusters
- clusters decay isotropically into two hadrons
- heavy clusters need to be treated separately



Indeed, the mass of color-singlet clusters is very small, and independent of the CM energy  $Q$ . Thus, the cluster model is relatively universal.

Light clusters decay into resonances & stable hadrons with  $\approx$  flat phase-space distribution. Heavy hadron production is thus suppressed.

However, long tail to high cluster mass values.



Heavy clusters undergo fission to lighter clusters ( $\rightarrow$  similar to string breaking)

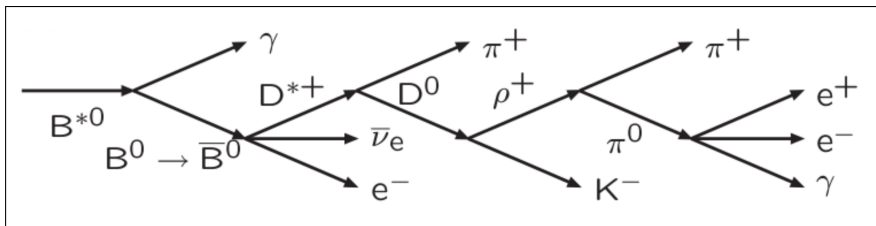
$\approx 15\%$  of primary clusters split

$\approx 50\%$  of hadrons emerge from split clusters



We are now approaching the final steps in the event generation chain.

Hadronization models often produce excited hadrons, which will decay within typical detectors. For example:



Note that some of these decays will leave displaced vertices, which may be important to “tag” heavy jets.

Majority of particles will be produced here; comprehensive machinery very important:

- Implement as many hadronic matrix elements as possible, especially for  $\tau$
- Include as many QED effects as possible
- Use PDG decay tables for rest. If incomplete, be creative.

Let us end on “If incomplete, be creative”.

## Summary of the lectures: Event generators are not magic.

Monte Carlo Event Generators use inversion and rejection sampling algorithms to produce events.

Events are pseudo-data that looks and feels very similar to real data.

Sophisticated pert. calculations used to predict inclusive x-sections, parton showers + multiple interactions to distribute these over many-parton states, using best insights into all-order QFT.

The parton  $\rightarrow$  hadron conversion is based both on perturbative and non-perturbative insights.

This level of detail does, however, come with a large number of parameters.

