QCD and jets
Lecture 3: Jets

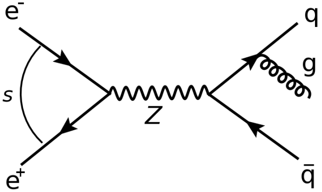
Sebastian Sapeta

CERN

Ecole Internationale de Physique Subatomique, Lyon, 20-25 October, 2014

Jets in e^+e^-: theoretical perspective

Real correction: $e^+e^- \to q\bar{q}g$



$$x_1 = \frac{2E_q}{\sqrt{s}} \qquad x_2 = \frac{2E_{\bar{q}}}{\sqrt{s}}$$

$$0 \le x_1, x_2 \le 1$$

$$x_1 + x_2 \ge 1$$

- x_1, x_2: fractions of energies carried by the quark and the anti-quark
- Differential cross section

$$\frac{d^2\sigma_{q\bar{q}g}}{dx_1 dx_2} = \frac{4\pi\alpha_{em}^2}{3s} 3e_q^2 C_F \frac{\alpha_s}{2\pi} \frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}$$

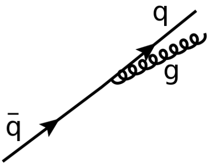
- x_is characterize the emitted gluon: $1 - x_{1,2} = x_{2,1} E_g (1 - \cos\theta_{2,1g})/\sqrt{s}$

$$\left.\begin{array}{ll} x_1 \to 1 & \text{when } \theta_{2g} \to 0 \\ x_2 \to 1 & \text{when } \theta_{1g} \to 0 \end{array}\right\} \text{collinear limit}$$

$$x_1, x_2 \to 1 \quad \text{when } E_g \to 0 \qquad \text{soft limit}$$

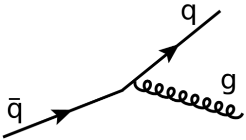
Jets in e^+e^-: theoretical perspective

- small-angle emission



$\hookrightarrow$ hitting collinear singularity: large correction

- large-angle emission

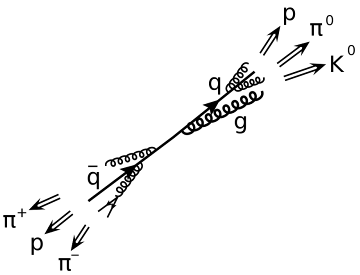


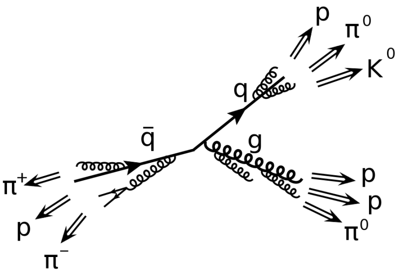
$\hookrightarrow$ away from collinear singularity: moderate correction

- Collinear singularity is at the end cancelled by the virtual correction but most of the emissions will anyway be soft and collinear leading to events with **two** highly collimated streams of particles.
- Some fraction of gluon emissions will however be large-angle and hard. That will produce events with **three** well separated objects.

Jets in e^+e^-: theoretical perspective

Reality is of course more complex: *further emissions* , *hadronization*.



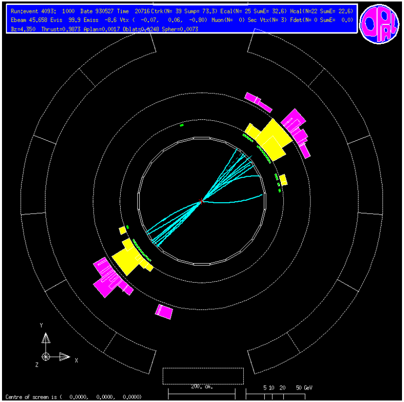


However, these two events remain clearly different and there is a well known reason for that: in one case the gluon is emitted off the $q\bar{q}$ pair at a small, and in the other case, at a large angle.

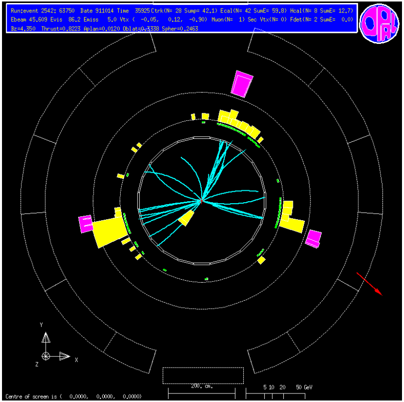
Jets in e^+e^-: experimental perspective

And it indeed looks like this in the experiment (OPAL at LEP1):

2-jet event



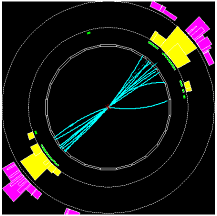
3-jet event



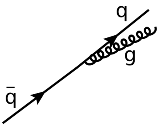
Jets in e^+e^-: connecting theory and experiment

We kind of see the correspondence:

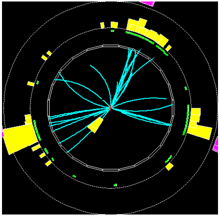
2-jet event



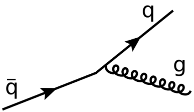
$\longleftrightarrow$



3-jet event



$\longleftrightarrow$



But we really need a quantitative measure that would allow us to classify events as 2- or 3-jets. Both in theoretical calculations and in experiment!
This is where a jet definition is needed.

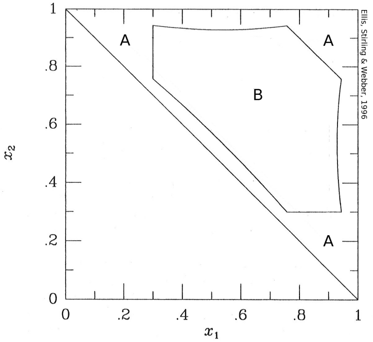
Jets in e^+e^-: Sterman-Weinberg definition

A first such definition was proposed by Sterman and Weinberg in 1977:

A final state is classified as a 2-jet event if at least a fraction $1-\epsilon$ of total available energy is contained in a pair of cones of half-angle δ.

$$\sigma_{\text{2-jets}} = \int_A dx_1 dx_2 \left(\frac{d^2\sigma_{q\bar{q}}^{\text{Born+Virt}}}{dx_1 dx_2} + \frac{d^2\sigma_{q\bar{q}g}}{dx_1 dx_2} \right)$$

$$\sigma_{\text{3-jets}} = \int_B dx_1 dx_2 \frac{d^2\sigma_{q\bar{q}g}}{dx_1 dx_2}$$



Fraction of 3- and 2-jet events:

$$f_3 \simeq \frac{g_s^2}{3\pi^2}\left(3\ln\delta + 4\ln\delta\ln 2\epsilon + \frac{\pi^2}{3} - \frac{7}{4}\right)$$

$$f_2 \simeq 1 - f_3$$

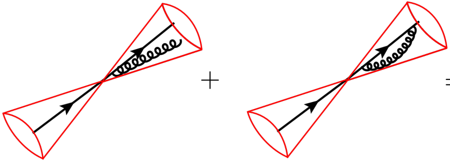
Infrared and collinear safety

Jet definitions are useful only if they are infrared and collinear (IRC) safe:

- emission of a soft or collinear gluon cannot change the set of jets

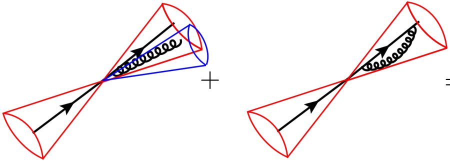
Why is this so important? Let's take a collinear emission off a quark

- IRC *safe* algorithm



$$= \frac{1}{\epsilon} f_X^{(2)} - \frac{1}{\epsilon} f_X^{(2)} = \text{finite}$$

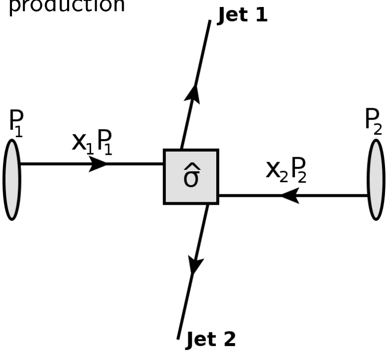
- IRC *unsafe* algorithm



$$= \frac{1}{\epsilon} f_X^{(3)} - \frac{1}{\epsilon} f_X^{(2)} = \infty$$

Jets at hadron colliders: factorization

Let's focus on dijet production



$$\sigma_{pp\to 2\text{jets}} = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij} \left(\alpha_s(\mu_R^2), \frac{Q^2}{\mu_F^2}, \frac{Q^2}{\mu_R^2} \right)$$

where, for dijets

$$\hat{\sigma}_{ij} = \alpha_s^2 \sum_{n=0}^{\infty} \alpha_s^n \hat{\sigma}_{ij}^{(n)}$$

and $n = 0$: LO, $n = 1$: NLO, $n = 2$: NNLO,...

Jets at hadron colliders: factorization

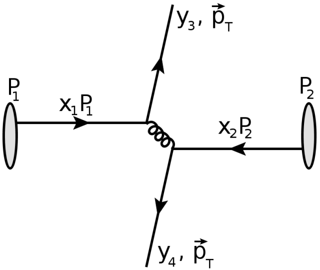
$$\sigma_{pp\to 2\text{jets}} = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij} \left(\alpha_s(\mu_R^2), \frac{Q^2}{\mu_F^2}, \frac{Q^2}{\mu_R^2} \right)$$

- at leading order, $\hat{\sigma}_{ij}$ is just a $2 \to 2$ partonic cross section and it is μ_F-independent
- at higher orders, $\hat{\sigma}_{ij}$ consists of the partonic cross section with subtracted collinear (long-distance) pieces, those pieces are transferred to PDFs – exactly like for Drell-Yan and DIS!
  - that you can do it is a subject of the factorization theorem for hadron-hadron collisions and it is proved to all orders [Collins & Soper '87]
  - after factorization, $\hat{\sigma}_{ij}$ is just a short-distance cross section; all the long-distance interactions are contained in PDFs; μ_F is the scale that separates the short- and the long-distance physics

Inclusive dijet production at leading order

- at LO jet = parton

$$\vec{p}_T = (p_x, p_y)$$
$$x_1 = \frac{p_T}{\sqrt{s}} \left(e^{y_3} + e^{y_4}\right)$$
$$x_2 = \frac{p_T}{\sqrt{s}} \left(e^{-y_3} + e^{-y_4}\right)$$



Differential cross section:

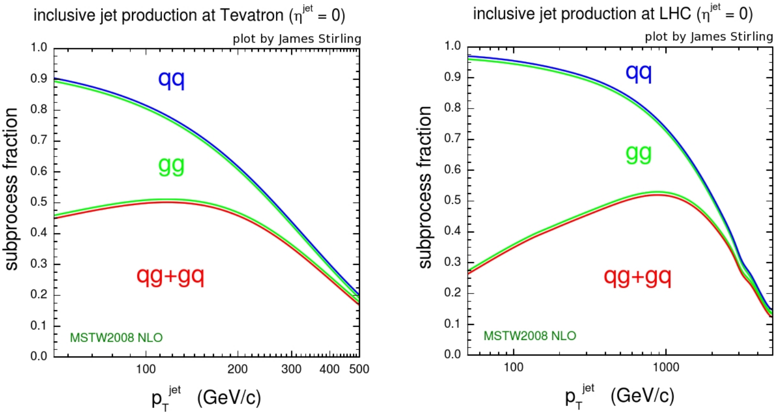
$$\frac{\sigma^{pp\to 2\text{jets}+X}}{dy_3 dy_4 dp_T^2} = \frac{1}{16\pi x_1 x_2 s^2} \sum_{i,j,k=q,\bar{q},g} f_i(x_1,\mu^2) f_j(x_2,\mu^2) \overline{\sum} |\mathcal{M}_{ij\to kl}|^2 \frac{1}{1+\delta_{kl}}$$

Contributing sub-processes:

$$qq \to qq \qquad qg \to qg \qquad gg \to gg$$

(and the same with the replacement $q \to \bar{q}$)

Contributions from different subprocess



- qq contribution is large at high p_T since that requires taking large $x_{1,2}$ from the proton and the quark PDFs have maximum at large x
- gg contribution is large at small p_T since that corresponds to probing the proton at low x and the gluon distribution is much larger than the quark distribution at low x

Modern jet algorithms for hadron colliders

Cone algorithms

- top-down approach
- historically the first: Sterman-Weinberg definition is of that type
- modern example: SISCone

Sequential recombination algorithms

- bottom-up approach
- repeatedly combine particles according to some distance measure
- modern examples: k_t, Cambridge/Aachen, anti-k_t

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \Delta R_{ij}^2 / R^2 \qquad d_{iB} = p_{ti}^{2p}$$

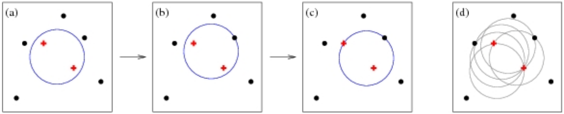
where $p = 1$ for k_t, $p = 0$ for CA, and $p = -1$ for anti-k_t

Common features of the above algorithms:

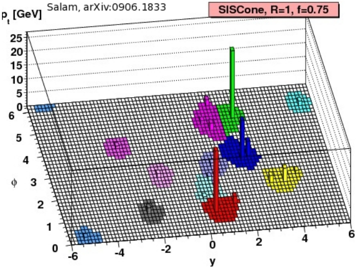
- infrared and collinear safety
- all final-state particles are included in the jets

SISCone algorithm [Salam, Soyez '07]

1. Find all distinct enclosures of radius R (distinct means that they have a different point content)



2. Test stability of each cone. This gives a set of protojets. Some of them may be overlapping.
3. Run split-merge procedure
   - order protojets according to scalar sum of transverse momenta $\tilde{p}_t$
   - for two hardest jets i and j with $p_{ti} > p_{tj}$: if shared $\tilde{p}_t > f\,\tilde{p}_{ti}$, where f is a parameter, merge the two protojets



   - otherwise split the protojets: particles assigned to closer jet
   - remove jets i and j from the list and repeat the procedure

Time required for clustering of N particles: $N^2 \ln N$.

Sequential recombination algorithms

1. Compute distances between particles for all particle pairs:

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p})\Delta R_{ij}^2/R^2 \,,$$

and the particle beam-distances for all particles:

$$d_{iB} = p_{ti}^{2p} \,.$$

- R is a jet radius
- ΔR_{ij} is a distance between the particles in the $y - \phi$ plane

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

2. Find smallest d_{ij} and d_{iB}:
   - if $d_{ij} < d_{iB}$, recombine the two particles and add the particle ij to the list of particles
   - if $d_{iB} < d_{ij}$, call i a jet and removed from the list of particles
3. Repeat the whole procedure until there is no particle left.

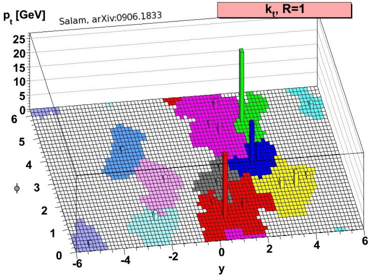
k_t algorithm [Catani, Dokshitzer, Webber, Seymour, Ellis, Soper '92-'93]

Distance measure:

$$d_{ij} = \min(p_{ti}^2, p_{tj}^2)\Delta R_{ij}^2/R^2 \qquad\qquad d_{iB} = p_{ti}^2$$

($p = 1$ case of the general formula)

- clustering starts from low-p_t objects that accumulate stuff around them
- for that reason, k_t algorithm produces jets that are very irregular in the $y - \phi$ plane
- well defined substructure



Time required for clustering of N particles: $N \ln N$.

Cambridge/Aachen algorithm

[Dokshitzer, Ledder, Moretti, Webber '97; Wobish, Wengler '99]

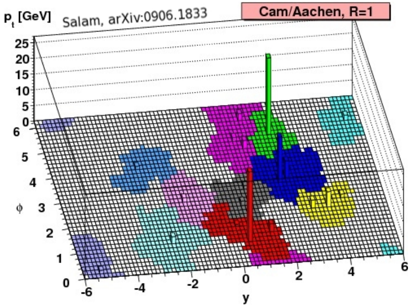
Distance measure:

$$d_{ij} = \Delta R_{ij}^2 / R^2 \qquad\qquad d_{iB} = 1$$

($p = 0$ case of the general formula)

- The measure is just a geometric distance in the $y - \phi$ plane.

- clustering insensitive to particle p_ts
- jets built up by merging particles closest in the $y - \phi$ plane
- well defined substructure



Time required for clustering of N particles: $N \ln N$.

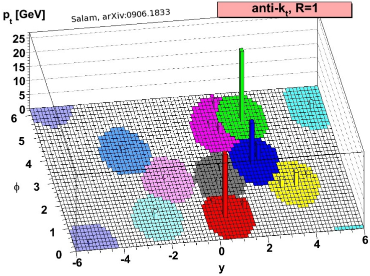
anti-k_t algorithm [Cacciari, Salam, Soyez '08]

Distance measure:

$$d_{ij} = \min(p_{ti}^{-2}, p_{tj}^{-2})\Delta R_{ij}^2/R^2 \qquad d_{iB} = p_{ti}^{-2}$$

($p = -1$ case of the general formula)

- Measure similar to the k_t alg. but with the replacement $p_{ti} \to \frac{1}{p_{ti}}$
- clustering starts from high-p_t objects that accumulate softer stuff around them
- clustering stops when there is nothing within radius R around the hard center: that gives jets which are very regular in the $y - \phi$ plane
- no well defined substructure



Time required for clustering of N particles: $N^{3/2}$.

FastJet

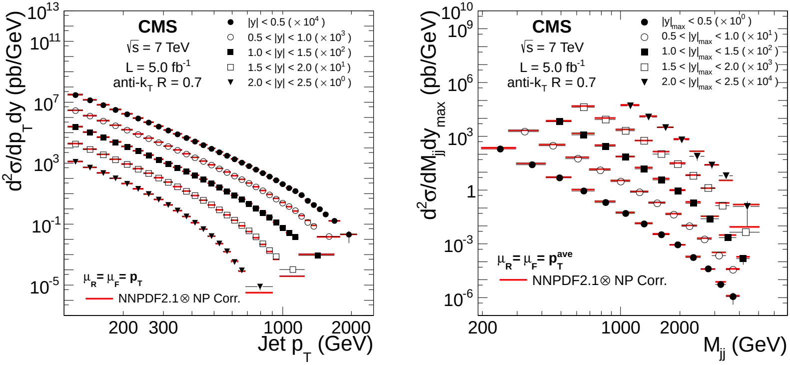
k_t, CA, anti-k_t, SISCone and many older algorithms are implemented in the FastJet package [Cacciari, Salam, Soyez]: http://fastjet.fr

Example

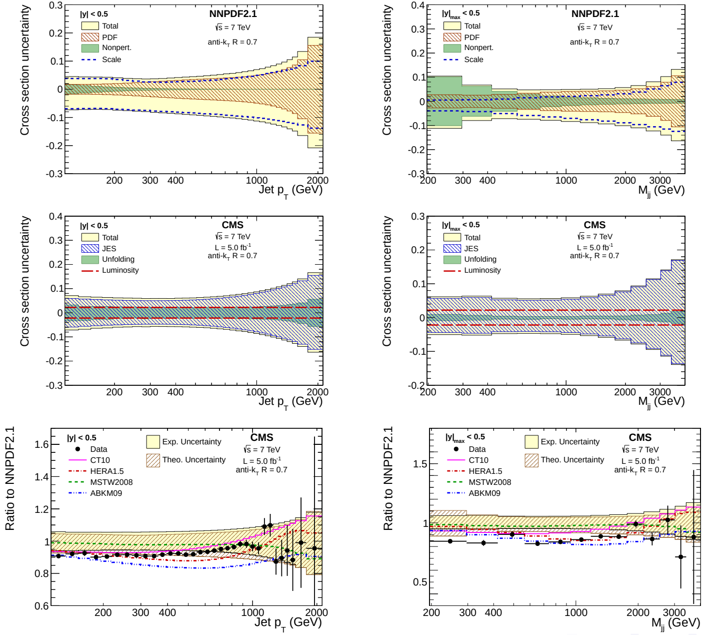
```
vector<PseudoJet> particles;
// an event with three particles:  px py pz E
particles.push_back( PseudoJet( 99.0, 0.1, 0, 100.0) );
particles.push_back( PseudoJet( 4.0, -0.1, 0, 5.0) );
particles.push_back( PseudoJet( -99.0, 0, 0, 99.0) );
// choose a jet definition
double R = 0.7;
JetDefinition jet_def(antikt_algorithm, R);
// run the clustering, extract the jets
ClusterSequence cs(particles, jet_def);
vector<PseudoJet> jets = sorted_by_pt(cs.inclusive_jets());
// print some info about the jets
cout << jet_def.description() << endl;
for (unsigned i = 0; i < jets.size(); i++) {
  cout << "jet " << i << ":  "<< jets[i].pt() << " "
       << jets[i].rap() << " " << jets[i].phi() << endl;}
```

Dijets at NLO vs data

[CMS, Phys. Rev. D 87 (2013) 11, 112002]

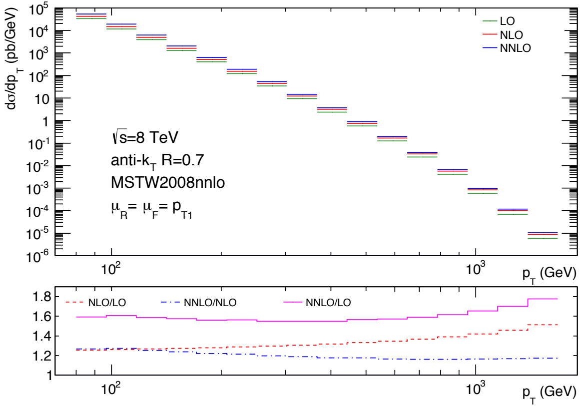


Dijets at NLO vs data



Dijets at NNLO

- only gg channel available so far [Currie, Gehrmann-De Ridder, Glover, Pires '13]



- Surprisingly large K-factor! Will it survive when all channels are added? We need to wait to see that...

Large K factors

Fixed order expansion

$$\sigma = \underbrace{\sigma_0}_{\text{LO}} + \underbrace{\alpha_s \sigma_1}_{\text{NLO}} + \underbrace{\alpha_s^2 \sigma_2}_{\text{NNLO}} + \cdots$$

Naively, one expects that

- $\sigma_i \simeq 1$ and $\alpha_s \ll 1$, hence the series should be nicely convergent
- μ_F and μ_R variation gives the uncertainty at each order n and the result at order $n+1$ should stay within this uncertainty

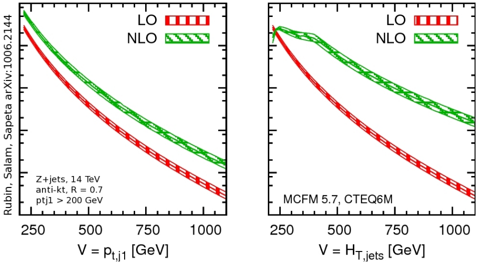
Like, for example, in the case of $e^+e^- \to$ hadrons:

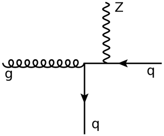
$$\begin{aligned} R(Q) &\equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} \\ &= 3\sum_q e_q^2 \left[1 + \frac{\alpha_s(Q)}{\pi} + 1.4\left(\frac{\alpha_s(Q)}{\pi}\right)^2 + \cdots\right] \\ &\simeq 3\sum_q e_q^2 \left[1 + 0.06 + 0.006 + \cdots\right] \end{aligned}$$

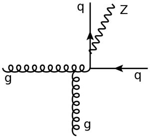
where we took 5 massless quarks, $Q = \sqrt{s} \ll m_Z$ and $\alpha_s(10\,\text{GeV}) \simeq 0.2$

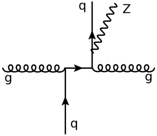
Large K factors

However, the situation often looks like this:







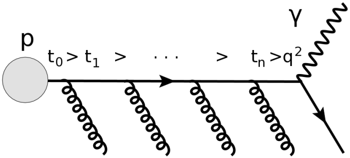


Origins:

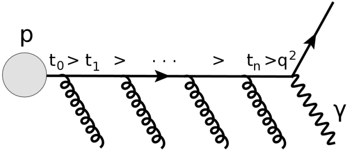
- new topologies
- new channels
- genuine loop corrections
- threshold logs
- and sometimes we just do not know

Beyond fixed order: types of branchings

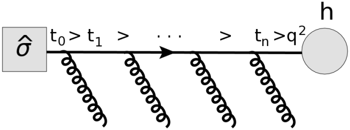
DIS
space-like
t-channel
initial-state



DY
space-like
t-channel
initial-state



jets
time-like
s-channel
final-state



Jets beyond fixed order

Time-like emissions, just like the space-like emissions, are enhanced at small-angles. Branching probability in the collinear limit reads:

$$d\sigma_{n+1} = d\sigma_n \frac{d\theta}{\theta} dz \frac{\alpha_s}{2\pi} \hat{P}(z)\,,$$

where θ is the angle between two partons created in the splitting.

- A parton created in hard process will iteratively emit real partons until its virtuality reaches the scale of $\mathcal{O}\left(1\,\mathrm{GeV}\right)$ when it hadronizes.

Evolution of the parent parton is described by the same DGLAP equation that we introduced in the context of DIS

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

where

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

is the no-emission probability: the Sudakov form factor.

Monte Carlo method

Introduction of the Sudakov form factor allows for construction of the parton shower

$$\Delta(t_1, t_2) \equiv \exp\left[-\int_{t_2}^{t_1} \frac{dt'}{t'} \int_{z_0}^{1-z_0} dz \frac{\alpha_s}{2\pi} \hat{P}(z)\right]$$

- $\hat{P}(z)$: unregularized splitting function; z_0: infrared cut-off

The algorithm (time-like branching):

1. Start from the parent parton at (t_1, x_1)
2. Generate the first emission at $t_2 < t_1$ by solving the equations

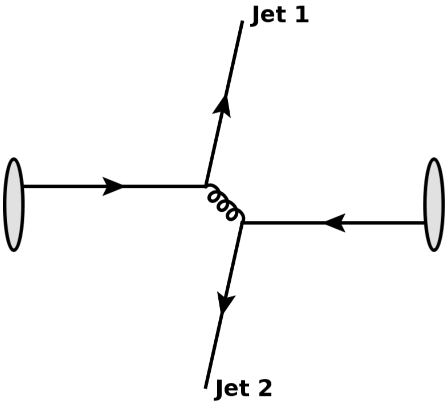
$$\Delta(t_1, t_2) = \mathcal{R} \qquad \text{and} \qquad \int_{z_0}^{x_2/x_1} dz \frac{\alpha_s}{2\pi} \hat{P}(z) = \mathcal{R}' \int_{z_0}^{1-z_0} dz \frac{\alpha_s}{2\pi} \hat{P}(z)$$

   where $\mathcal{R}, \mathcal{R}' \in [0, 1]$ are random numbers.
3. If $t_2 > t_{\min}$, repeat the procedure, otherwise quit the loop.

- generates resolvable emissions $(z_0 < z < 1 - z_0)$ up to the scale $t_{\min}$

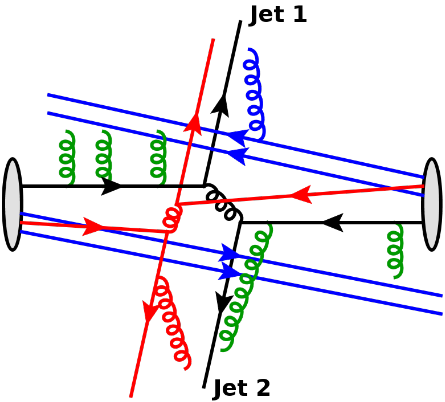
Complications at hadron colliders

Modeling jets as partons from hard $2 \to 2$ collision is somewhat simplistic



Complications at hadron colliders

Real proton-proton collisions give us

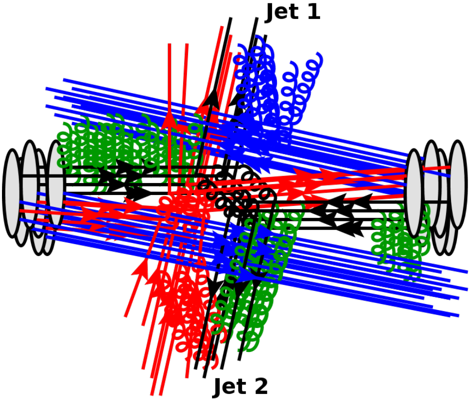


- proton remnants
- initial state radiation
- multi-parton interactions

⇒ all this goes under the name of the underlying event

Complications at hadron colliders

And real proton-proton collisions at the LHC looks even worse



- many simultaneous *pp* collisions per bunch crossing: pileup

Contamination from UE/PU

For a jet of radius R, the following amount of its p_t and mass comes from the underlying event/pileup (on average over many events):

$$\langle \delta p_t \rangle \simeq p_{t,\mathrm{jet}}\, \rho\, \pi R^2 \,, \qquad \langle \delta m^2 \rangle \simeq p_{t,\mathrm{jet}}\, \rho\, \pi \frac{R^4}{2} \,,$$

where ρ is the average UE/PU momentum per unit area.

- Contamination from UE/PU increases with jet size.
  $\hookrightarrow$ going for very small jets does not solve the problem since one becomes sensitive to hadronization which goes like $\delta p_t \propto -\frac{1}{R}$
- Typical ρ values for the LHC:

  $$\rho_{\mathrm{UE}} \simeq 2\,\mathrm{GeV} \qquad \rho_{\mathrm{PU}} \simeq 10 - 20\,\mathrm{GeV}$$

  This has a significant effect on steeply falling spectra!
- For real events, one makes the replacement

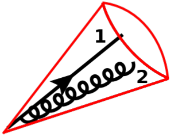
  $$\pi R^2 \to \text{jet area} \qquad \pi R^4/2 \to \text{jet mass area}$$

Determining ρ allows one to subtract the contamination from p_t and mass.

Jets are massive

...or at least those jets that contain more than one particle.

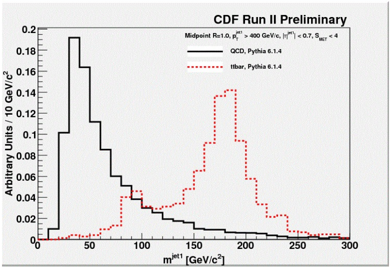
A small-radius jet consisting of two massless partons has the mass



$$m_J^2 \simeq p_{T1} p_{T2} \Delta_{12}^2 \,,$$

where $\Delta_{12}^2 = (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2$.

Mass spectra are very different for QCD jets (from splittings like $q \to qg$) and for jets from decays of heavy objects ($Z \to q\bar{q}$)



Jets substructure

Jets from algorithms like k_t or CA have a well defined substructure:

- one can interpret the clustering sequence as following the actual emission sequence backwards.

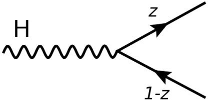
Analysis of jet substructure is powerful in signal/background separation.
Suppose that you are interested in the process

$$pp \to ZH \to Z\, b\bar{b}\,,$$

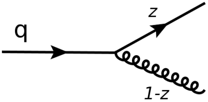
If you require very high p_t of the Z boson, say 200 GeV, the $b\bar{b}$ pair will end up it the same jet.

$$Z\, b\bar{b} \to Z\; +\; \text{jet}$$

But this has a huge background from ordinary $pp \to Z + \text{jet}$ process!
The substructure of jets looks however very different in those two cases:



$$P(z) \propto 1$$



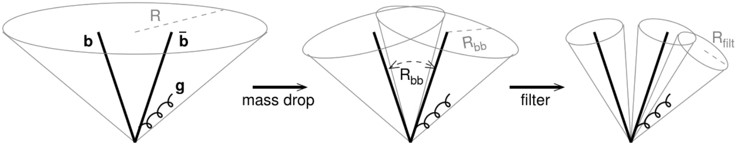
$$P(z) \propto \frac{1+z^2}{1-z}$$

Jets substructure

- Higgs decays symmetrically in z whereas QCD jets have strong enhancement for asymmetric share of energy $z \to 1$.

We can use this for improving signal/background ratio!

Example for ZH production: [Butterworth, Davison, Rubin and Salam '08]



1. Cluster with C/A, select events with $p_{T,Z} > 200\,\text{GeV}$ (boosted events).
2. Undo the last clustering $j \to j_1 j_2$ labeled such that $m_{j_1} > m_{j_2}$
3. If $m_{j_1} < \mu m_j$ and $\min(p_{j1}, p_{j2})/\max(p_{j1}, p_{j2}) > y_{\text{cut}}$ exit the loop
4. Redefine j to be equal to j_1 and go back to step 1.
5. Recluster subjects j_1 and j_2 with $R = R_{\text{filt}}$, take three hardest subjects and construct the mass.

Summary of lecture 3

- In QCD, the collinear gluon emissions are enhanced and the large-angle emissions are rare. Therefore most of the final state particles appear in the form of collimated bunches called jets.
- Jets are indeed observed in experiments and jets observables can also be computed theoretically.
- To a first approximation jets can be regarded as a quark or a gluon that took part directly in the hard scattering.
- To relate the jets of hadrons, registered by detectors, to the jets of partons, computed in pQCD, we need a robust jet definition. That allows for a meaningful comparison between theory and experiment.
- We discussed the modern jet algorithms: k_t, CA, anti-k_t & SISCone.
- At hadron colliders, one needs to address the question of jet contamination coming from the underlying event and pileup.
- Fixed order calculations for jet production processes are known only up to NLO so in order for realistic simulation of the final states we often resort to the parton shower algorithms, which give us many-particle final states using the collinear approximation.
- Jets have substructure whose analysis is a powerful tool for signal-background discrimination.

References

- **QCD and Collider Physics**
  R.K. Ellis, W.J. Stirling and B.R. Webber
  Cambridge University Press 1996
- **Quantum Chromodynamics**
  High Energy Experiments and Theory
  G. Dissertori, I. Knowles and M. Schmelling
  Oxford University Press 2003
- **Foundations of Perturbative QCD**
  J. Collins
  Cambridge University Press 2013
- **Handbook of Perturbative QCD**
  CTEQ Collaboration
  http://users.phys.psu.edu/~cteq/#Handbook
- **Towards Jetography**
  G. Salam
  Eur. Phys. J. C **67** (2010) 637, arXiv:0906.1833