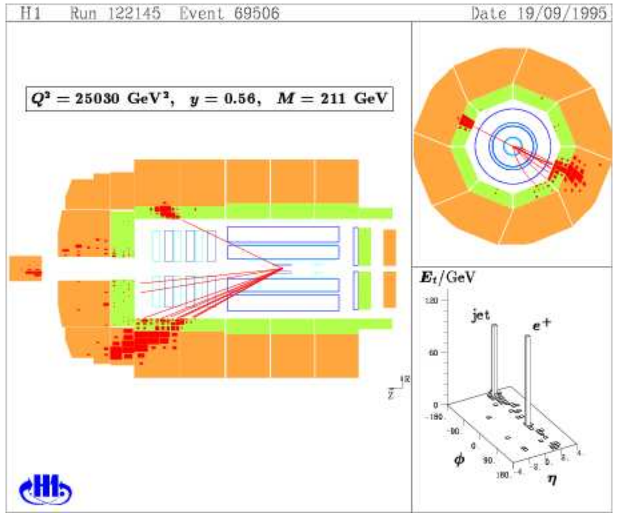
Jet algorithms

Sergei Chekanov (ANL)

28 Sep. - 4 Oct. 2002

14th Topical Conference on Hadron Collider Physics

- ⋆ What are jets and why they are interesting ?
- ⋆ Variables and reference frames.
- ⋆ Requirements on a jet algorithm.
- ⋆ Clustering algorithms: JADE vs Durham.
- ⋆ The cone algorithm.
- ⋆ The longitudinally invariant $\boldsymbol{k}_{\perp}$ algorithm.
- ⋆ Summary.



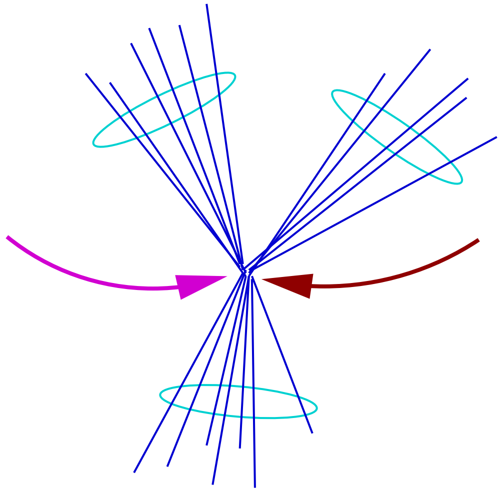
What is a jet ?

Jets - sprays of objects

1. Simplified characteristics of events

$$\{p_1, p_2, \ldots, p_n\} \rightarrow \{P_1, P_2, \ldots, P_N\}, \quad n >> N$$

2. Reduction of information helps to focus on main properties of the underlying dynamics
3. One jet - one parton assumption
4. Jets allow to study short-distance aspects of QCD (color dynamics, the strong coupling constant (α_s), reconstructions of heavy particles ($Z, W, t,$ etc.)

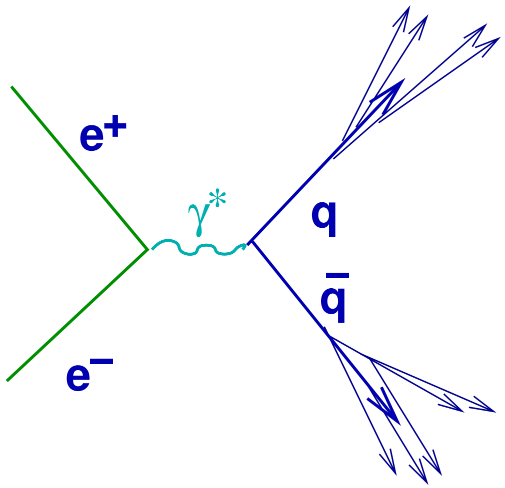


Variables for jet search

This depends on the process under study:

e^+e^- annihilation:

- ✔ centre-of-mass system $\longrightarrow$ Lab. frame
- ✔ invariance under rotations
- ✔ $\longrightarrow$use energies (E_i) and angles (θ_i) as a distance between two particles

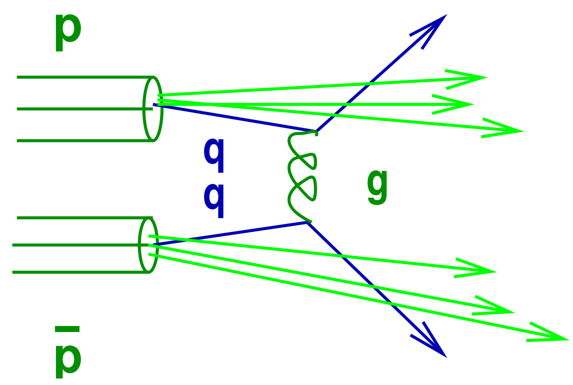


The logical choice is $d_{ij}^2 = 2E_iE_j(1 - \cos\theta_{ij})$ which equals to invariant mass of two particles (assuming $m_i = m_j = 0$)

Variables for jet search (cont'd)

$p\bar{p}$ and ep collisions:

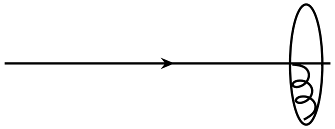
- ✔ initial-state system is not at rest: HCM/ Breit frame;
- ✔ the final-state system is boosted along the beam axis;
- ✔ angular separations θ_i are NOT invariant under boost;
- ✔ Variables which are invariant under boost and which allow to treat on equal footing all final-state hadronic systems are:
  - ✔ transverse energy ($E_\perp$)
  - ✔ azimuthal angle (ϕ)
  - ✔ pseudorapidity difference ($\Delta\eta$)

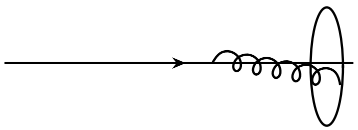


The distance defined as $R = \sqrt{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}$ is invariant under longitudinal boost and reflect BOTH the longitudinal and angular separations between two particles

Requirements on a jet algorithm

- ⋆ Measurable and applicable to any object (calorimeter cells, tracks, energy flow objects) used in experimental analyses
- ⋆ Calculable in theory:
  - – infrared and collinear safe: does not change if a soft particle is added or one particle splits into two collinear particles





- ⋆ Close correspondence with the parton direction
- ⋆ Suppression of soft processes (hadronisation) and beam remnants
- ⋆ simple to use

Requirements on a jet algorithm (cont'd)

The decision on which algorithm to use has to be based on understanding of the size of:

★ higher-order QCD contributions:

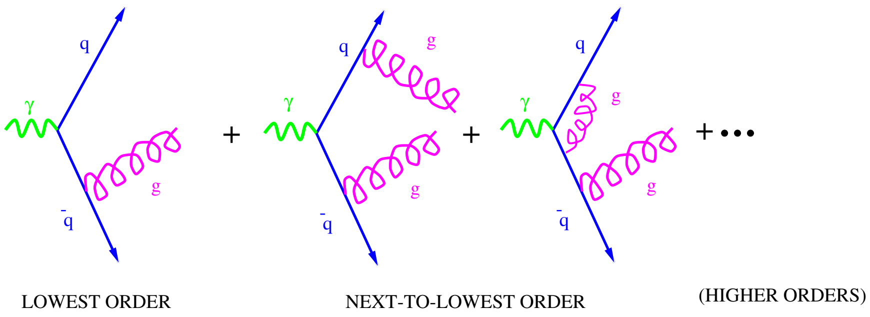
$A = A_1 \cdot \alpha_s(\mu_R) + A_2 \cdot \alpha_s^2(\mu_R) + \ldots higher - orders$

$\boxed{A \text{ does not depend on } \mu_R \text{ at all orders}}$

higher-order contributions can be estimated (without computing them) by investigating the renormalization scale dependence - μ_R

Typical choice: $E_\perp/2 < \mu_R < 2E_\perp$ if $\mu_R = E_\perp$

good jet algorithm should have a small dependence on μ_R!



★ hadronisation correction and hadronisation uncertainties

★ experimental uncertainties

Hadronisation corrections and uncertainties

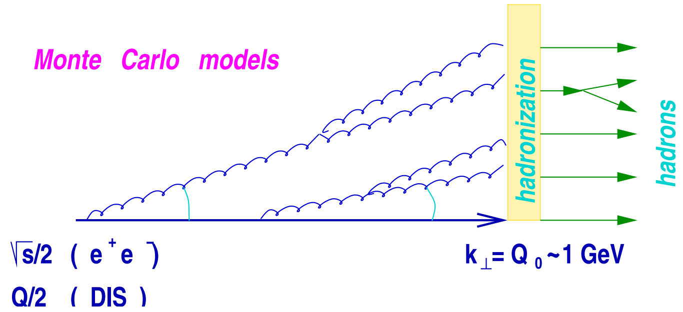
At low $\boldsymbol{E}_\perp$ of jets, the contribution from hadronisation can be significant.

At HERA, the following procedure is adopted:

Hadronisation correction:

$$\boldsymbol{C = \sigma^{MC}_{partons} / \sigma^{MC}_{hadrons}}$$

using available Monte Carlo models



Hadronisation Uncertainties: the hadronisation corrections should be computed for a few different Monte Carlo models (usually - HERWIG and Lund-string based)

Optimal jet algorithm should have $\boldsymbol{C \sim 1}$ and small differences between hadronisation corrections evaluated using different Monte-Carlo models

Recombination procedure: JADE

1. Distance measure $d_{ij} = 2E_iE_j(1 - \cos\theta_{ij})$ for each particle pair;
2. define a minimum separation d_{cut};
3. particles are merged into clusters with 4-momenta $p_i + p_j$ if $d_{ij} < d_{cut}$;
4. go to step (1) until all pairs of clusters have a separation above d_{cut} (note d_{ij}/E_{vis} is more often quoted). Whatever objects remain at this stage are called jets;

JADE is popular algorithm for e^+e^-physics and early days of HERA studies

The limitation of the JADE algorithm:

soft gluons (copiously radiated!) far apart can have small d_{ij} :

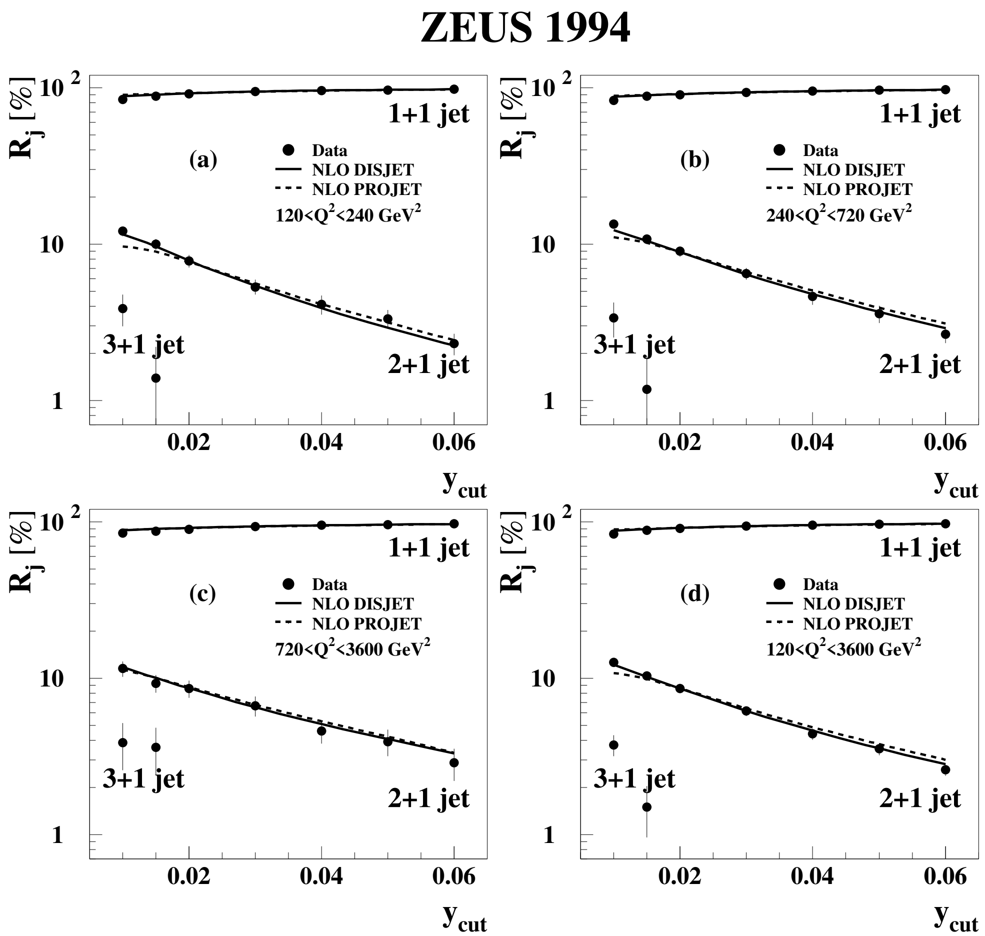
$$\theta >> 0, \quad E_i \sim E_j \sim 0$$

and thus be combined into a "phantom" $\longrightarrow$**sizable fixed-order QCD corrections**

For $\boldsymbol{ep}$, the JADE was modified to deal with the proton remnant: a pseudo-particle was inserted to take into account the missing longitudinal momenta

$\longrightarrow$helps to combine soft particles into "remnant jet"

Popular algorithm at HERA in the past



The $k_\perp$ (Durham) algorithm

replace $d_{ij}^2 = 2E_iE_j(1 - \cos\theta_{ij})$ by $d_{ij}^2 = 2\,\min(E_i^2, E_j^2)\,(1 - \cos\theta_{ij})$ (S.Catani et al (1991))

- ⋆ the measure is the transverse energy (not momenta!) for small angles:

$$d_{ij}^2 \simeq \min(E_i^2, E_j^2)\,\sin^2\theta_{ij} = k_\perp^2$$

- ⋆ the transverse momentum of the lower energy parton to a reference direction given by the high-energy parton

The $k_\perp$ algorithm has taken over from JADE as the standard jet finder

For the Durham distance measure soft gluons will be combined first with nearby higher-energy parton:

$\Longrightarrow$ **the stability of the fixed-order results is greatly improved!**

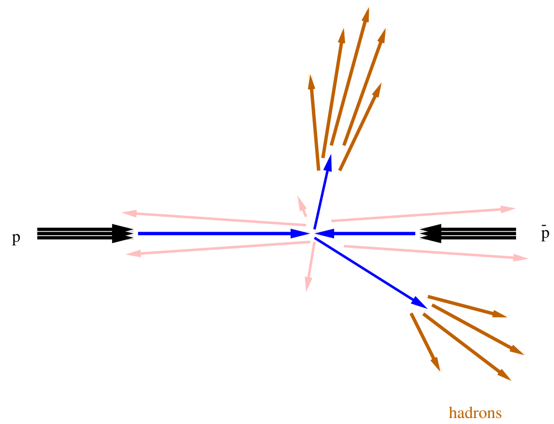
(rather similar e^+e^- algorithms: LUCLUS, GENEVA, CAMBRIDGE)

Differences between e^+e^- and hadron collisions

An algorithm for hadron-hadron collisions should naturally to take into account differences between e^+e^- and hadron-hadron collisions:

- ✔ Measured cross sections are different: In e^+e^- , entire event arises from initially small number of partons . Thus, exclusive cross sections describing the production of exactly N jets and nothing else. In hadron collisions - the practice has been to measure inclusive large p_T cross sections;
- ✔ Variables are different (see above);
- ✔ For hadron-hadron collisions, only a small fraction of the final-state hadrons are associated with high p_T jets, keeping the high p_T jets distinct from the debris in the beam jets.

- ⋆ Beam remnant jets have huge energies, but they have not undergone a hard scattering
- ⋆ Large momentum transfer - small distance (hard scattering)



The $k_\perp$ algorithm for DIS

the e^+e^- Durham algorithm was modified for DIS to minimize contamination from the hadron remnant (S.Catani, Yu.Dokshitzer, B.Webber, 1992):

★ Define a hard scale (usually E_T);

★ for every hadron k at angle θ_k with respect to the initial hadron direction, as well as for every hadron pair k and l with angle θ_{kl} between them, compute:

$$y_k = \frac{2(1-\cos\theta_k)}{E_T^2} E_k^2 \qquad y_{kl} = \frac{2(1-\cos\theta_{kl})}{E_T^2} \min(E_k^2, E_l^2)$$

★ take smallest value y_k, y_{kl} . If y_{kl} is the smallest value and $y_{kl} < 1$, combine momenta of 2 hadrons into a single cluster. If y_k is the smallest value and y_k- include this hadron in the beam jet;

★ repeat this procedure until all clusters have y_k, $y_{kl} > 1$;

★ one ends up with N number of hard jets and the beam jet $\longrightarrow$ set experimental cuts;

★ jet substructure for macro jets can be studied using a procedure similar to e^+e^-;

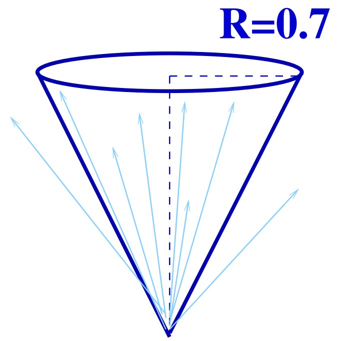
★ the algorithm was proposed for the Breit frame and measures **exclusive** jet cross sections, similar to e^+e^-

A SIMILAR PRESCRIPTION WAS PROPOSED FOR $p\overline{p}$ COLLISIONS, BUT THE $p\overline{p}$ COMMUNITY TOOK A DIFFERENT ROUTE...

Cone algorithm (The Snowmass Accord,1990)

- ⋆ particles should lie within a cone centered on the jet axis:

$$\sqrt{(\eta_n - \eta)^2 + (\phi_n - \phi)^2} < R$$



- ⋆ Jet momentum, pseudorapidity, and azimuth are defined as

$$E_T = \sum_{i \in cone} E_{Ti} \quad \eta = \frac{1}{E_T} \sum_{i \in cone} E_{Ti}\eta_i$$

$$\phi = \frac{1}{E_T} \sum_{i \in cone} E_{Ti}\phi_i$$

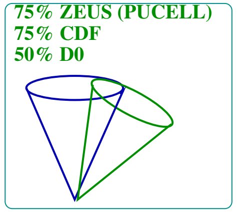
From the experimental point of view, it is was not a complete definition $\Longrightarrow$ the maximization process is not uniquely defined - need for seed directions (details vary from experiment to experiment)

$R = 0.7$ was suggested at the Snowmass

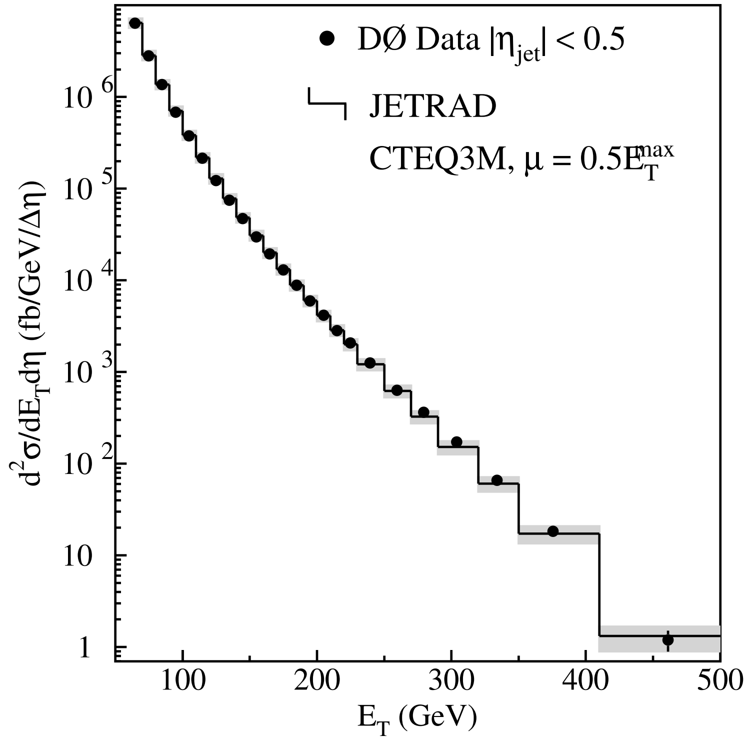
Cone algorithms (cont'd)

There are another problems with the cone algorithms:

1. Jets often overlap and share energy



an arbitrary procedure is needed to split/merge overlapping cones



2. ad-hoc parameter, $\boldsymbol{R_{sep}}$, is required to accommodate differences at parton and detector levels:

3. not infrared safe at the NNLO QCD! (W.Giele,W.Kilgore (1997) and M.Seymour (1998))

Need an additional stage of iterations!

Getting more and more complicated !

See also recent innovation suggested at the Workshop "Physics at Run II"- The Improved Legacy Cone Algorithm

Coming back to the iterative algorithm

all these problems can be solved using the iterative longitudinally invariant $k_\perp$ algorithm defined for **inclusive** cross sections, similar to the cone algorithm (S.Ellis and D.Soper, 1993):

- ★ Define for each protojet (hadron) $d_i = E^2_{T,i}$ and a parameter $R \sim 1$ analogous to the cone size ;
- ★ for each protojet pair compute:

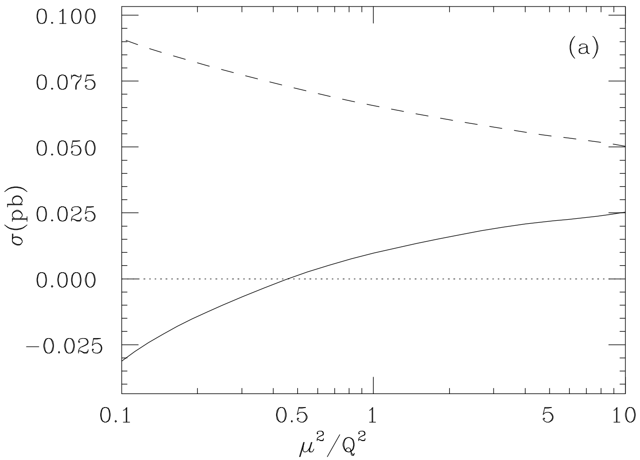
$$d_{ij} = \min(E^2_{T,i}, E^2_{T,j}) \frac{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}{R^2}$$

- ★ if the minimum of all $d_i, \; d_{ij}$ is d_{ij}, then the protojets are merged. If the minimum is d_i, the protojet is called jet and removed from the iterations.
- ★ gives a list of jets with successively larger $d_i = E^2_{T,i}$, then require certain criteria:
  - – jet pseudorapidities should be in some range;
  - – jet transverse energy $> E_0$ (the lower E_0, the larger exp. and. theor. uncertainties)

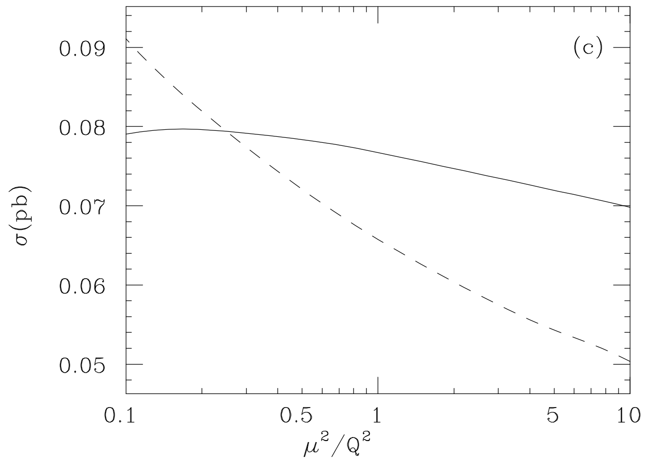
the algorithm is very similar to the generalized Durham algorithm, but maintains a similarity with the cone algorithm

from fixed-order perturbative QCD: $R_{cone} = R/1.35$

Comparisons: the cone vs $k_\perp$ algorithm



CDF cone algorithm



$k_\perp$algorithm

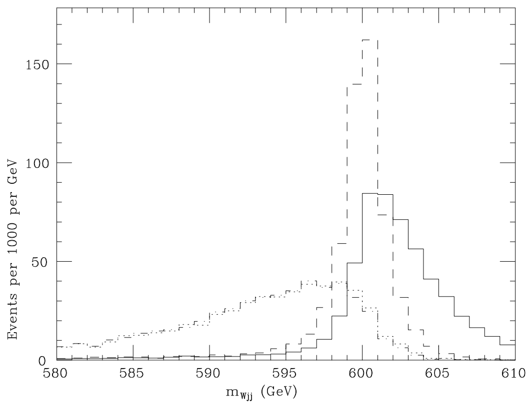
(LO - dashed lines , NLO - solid lines)

The 2-jet cross sections at high Q^2 DIS for LO (dashed lines) and NLO (solid lines)

(FROM B.PÖTTER AND M.SEYMOUR, J.PHYS. G25 (1999) 1473)

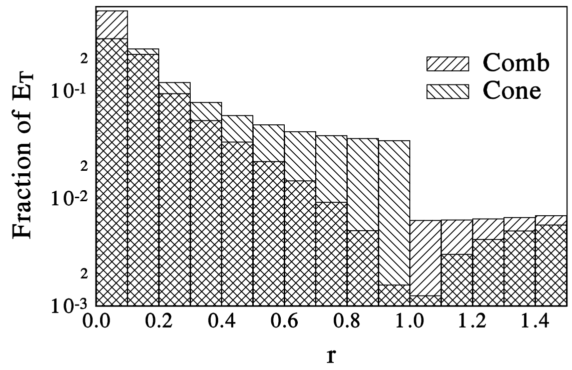
The CDF cone algorithm is very unstable for high Q2 DIS!

Comparisons: the cone vs $k_\perp$ algorithm



Reconstructed mass distribution of Higgs candidates using the cluster (solid and dashed lines) and the cone (dotted and dot-dashed lines) algorithms

(see M.Seymour, NP B421 (1994) 534)



Energy spreads within jets are very different:

- ⋆ The $\boldsymbol{k}_\perp$ pays more attention to the jet core;
- ⋆ The cone algorithm sucks too many soft gluons into edges

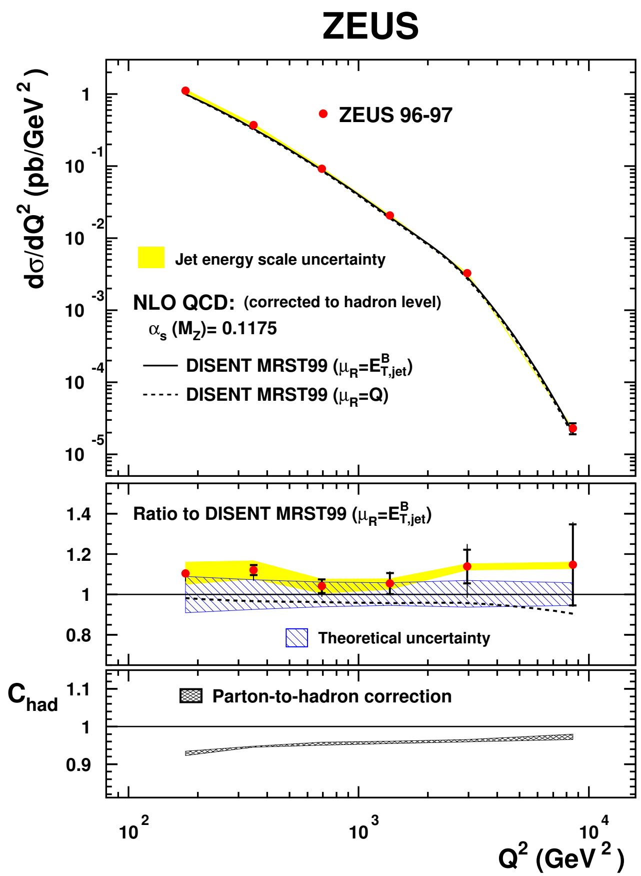
(see S.Ellis, D.Soper, PR D48 (1993) 3160)

The longitudinally invariant $k_\perp$ and cone algorithm

Performance:

- ✘ small higher-order QCD effects ($\mathbf{10-20\%}$)
- ✘ small hadronisation effects ($< \mathbf{10-20\%}$)
- ✘ small hadronisation uncertainties ($< \mathbf{3\%}$)
- ✘ small experimental uncertainties ($< \mathbf{3\%}$)
- ✘ resolution for $E_\perp > \mathbf{17}$GeV:
  - ✘ $\phi^{jet} \sim 2^0 \; \eta^{jet} \sim 0.04 \; E_\perp \sim 5\%$
- ✘ can be used for DIS and photoproduction jets
- ✘ sub-jet multiplicities as a function of y_{cut} can be studied reapplying the k_T algorithm
- ✘ allows a precise determination of α_s in DIS (see O.Gonzalez's presentation)

the main choice at HERA for DIS and photoproduction jets



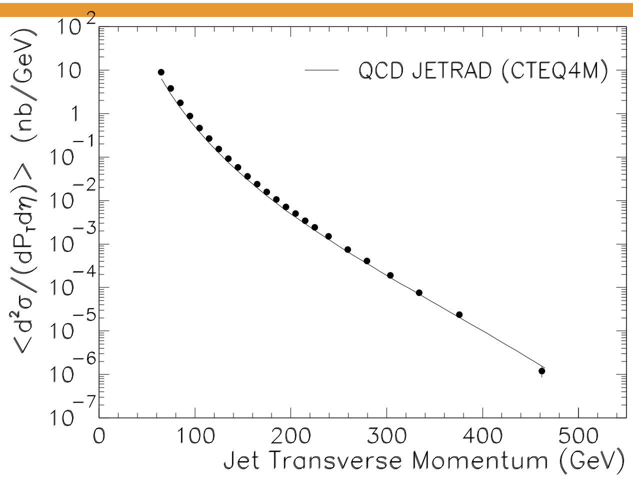
Numbers of published papers by H1 and ZEUS Collaborations

Blue - H1 Red -ZEUS

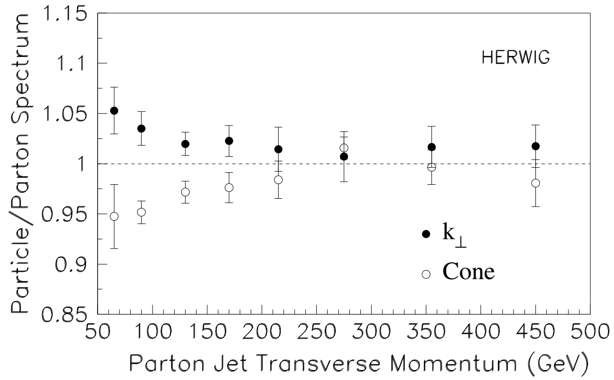
year of publication	JADE	Cone (CDF like)	$k_\perp$ exclusive	$k_\perp$ long. invariant.
1995	1,2	1,2	1	
1996		1		
1997		1,1		1
1998	2	3,2		2,2
1999		1,1	1,1	1
2000		1	2,1	1,2
2001		1		3,3
2002				2,2

The longitudinally invariant $k_\perp$ algorithm

The first D0 measurement (2001) of the inclusive cross section using the $k_\perp$ algorithm:



- ⋆ $R = 1$ is used
- ⋆ trigger selection is based on the cone algorithm with $R_{cone} = 0.7$



- ⋆ Hadron-to-parton correction is different for the the $k_\perp$ algorithm and cone algorithm
- ⋆ Energy loss outside the cone ?
- ⋆ .. or proton remnant contributions to the $k_\perp$ jets ?

Summary and future developments

- ✔ Experiments should abandon cone-based algorithm and use the versions of the $\boldsymbol{k}_\perp$ algorithm; if for some reason this is not possible the cone algorithm should be improved;

- ✔ All algorithms discussed above have two conceptual problems: it is hard to study them theoretically and they have instabilities resulting from their discontinuity (not stable with respect to small changes of input data);

  ⋆ $\longrightarrow$hadronic energy flow based on $\boldsymbol{C}$ -algebra is needed (F.Tkachov, 1995)?

- ✔ Future development of jet algorithms depends on the progress in QCD (understanding of higher-order QCD correction and hadronisation)