JET FINDING ALGORITHMS AT TEVATRON*

B. Andrieu

LPNHE, Universités Paris VI/VII, IN2P3-CNRS, Paris, France

(Received November 8, 2004)

A description of the jet finding algorithms used in the CDF and DØ experiments in Run II is given.

PACS numbers: 12.38.Qk

1. Introduction

"Hard" (*i.e.* high transverse energy) processes in hadron-hadron collisions are understood as footprints of interactions between two partons of the incoming hadrons. In QCD, partons are identified to the quarks and the gluons and cross sections can be calculated in perturbative QCD (pQCD). Cross sections can also be evaluated with the help of Monte Carlo event generators, *e.g.* for processes where many partons are emitted, by applying parton showering to processes of known cross section. Due to color confinement, the hard outgoing partons fragment into many uncolored particles. The partons which do not participate in the hard scattering can undergo soft processes (the underlying event) which add energy to the hard process. Even if the fragmentation and underlying event are not described in QCD yet, the existence of the observed "jets" (*i.e.* high energy flows of collimated particles) can be derived from QCD [1]. Jet algorithms can be used to reconstruct jets at the parton, particle and detector levels. Provided jets are defined as to avoid collinear and infrared singularities, partonic pQCD calculations can be regarded as predictions for jets. If measurements and pQCD predictions are corrected to the particle level, they can then be compared in a meaningful and consistent way. After recalling general features of jet algorithms in the Tevatron context in Section 2, the two main jet finding algorithms used in the CDF and DØ experiments at Run II will be described: the Cone algorithm in Section 3 and the $k_\perp$ algorithm in Section 4. Section 5 will provide a summary and suggestions for the future.

* Presented on behalf of the DØ Collaboration at the XXXIV International Symposium on Multiparticle Dynamics, Sonoma County, California, USA, July 26–August 1, 2004.

2. General features

2.1. Jet algorithm definition

Jets are formed as exclusive subsets of nearby items[1], where the distance used can be either angular (*e.g.* $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$ for cone algorithms, where η is the pseudorapidity and φ the azimuthal angle) or also proportional to energy (*e.g.* relative transverse momentum for $k_\perp$ algorithms). The rules followed to group items constitute the jet algorithm definition. Each jet is eventually defined as the combination (Section 2.3), performed according to a recombination scheme[2], of all its items.

2.2. Requirements for jet algorithms at Tevatron

In order to ensure a valid and precise comparison between jet cross sections measured experimentally (detector level) and predicted from QCD calculations (parton level), both corrected to the particle level, jet algorithms should act in the same way at parton, particle or detector level (order independence) and provide, for each event, a result which is independent of the emission of very low energy partons or particles (infrared safety) and of the collinear splitting of partons or sharing of energy between adjacent towers (collinear safety). In addition, kinematic variables used to describe jets should exhibit boundaries insensitive to the details of the final state (boundary stability). At Tevatron, Lorentz invariance under longitudinal (*i.e.* along the beam axis) boosts is also required. From the experimental point of view, desired features [4] are: independence of detector characteristics, minimal sensitivity to experimental (noise, luminosity,...) conditions, maximal jet finding efficiency in a minimal CPU time, minimal effect of resolution smearing and angle bias, minimal hadronisation and detector corrections and reliable jet energy calibration. The algorithms should also be fully specified and straightforward to implement and should give, as far as possible, results compatible with those obtained using Run I algorithms.

2.3. Combination of items

In Run I, the various recombination schemes used were based on E_{T}-weighted means, as inspired by the original Snowmass scheme [5]. This preserves longitudinal boost invariance in a very simple way, but the jet E_{T} is not a true energy variable and does not respect boundary stability. For

[1] The generic term "items" will be used to designate either of the three possible inputs to jet algorithms: partons, particles or detector objects. When the text refers to calorimeter objects only, the term "towers" is used instead.

[2] A recombination scheme, which may be different from the one used for the final computation of jets, is also used in various steps of the jet algorithm itself.

Run II, the E-scheme, which uses the full 4-vector information, is chosen instead [4]. The usual kinematic variables of the combined object are then calculated using its 4-momentum, in particular the transverse momentum p_{T}, which respects the required boundary stability. Provided the distance used to associate items to jets is invariant under longitudinal boosts, jet algorithms based on this scheme respect the same property.

2.4. Experimental facts

The CDF [2] and DØ [3] detectors are described in detail elsewhere. Here only the minimal information on the main subdetector used for jet reconstruction in each experiment, *i.e.* the calorimeter, is given. The CDF calorimeter uses lead or iron as absorber and scintillator as active medium. It consists of ∼1500 towers, each divided into an electromagnetic (e.m.) and a hadronic (had.) section, of minimal size 0.11×0.13 in (η, ϕ). Each tower section (e.m. or had.) is attributed a massless 4-momentum of energy the energy deposited in that section and of direction the 3-vector joining the interaction vertex to the center of that section. The 4-momentum of a tower is obtained by combining the 4-momenta of its e.m. and had. parts according to the E-scheme (Section 2.3). Only towers with $p_{\mathrm{T}} > 100$ MeV are used in jet algorithms. Similarly, the DØ calorimeter uses uranium and liquid argon and comprises 45000 cells combined, according to the E-scheme too, in ∼5000 towers of minimal size 0.1×0.1 in (η, ϕ).

3. The Midpoint or Run II Cone algorithm

Cone jet algorithms group items in cones of origin the interaction point and of fixed angular[3] radius R_{cone}. A proto-jet corresponds to a "stable" cone, *i.e.* whose axis coincides with the direction of the jet. Since stable cones cannot be found analytically, an iterative procedure is used. Proto-jet candidates (cones) are first formed around axes in given starting positions and each proto-jet direction is calculated from the combination of all items in its cone. The process is then iterated, replacing the cone axis by the proto-jet direction, unless or until the two coincide.

For computing time reasons, proto-jets cannot be searched starting from each position corresponding to every point on a (η,φ) grid, which should be fine enough to allow all stable cones to be found. Instead, only "seeds", *i.e.* items with a sufficient p_{T}, are used. In DØ, due to the high number of seeds per event in data, a preclustering (Section 3.1) is also needed to diminish the number of seeds by using preclusters instead of towers as seeds.

[3] In Run II, to be consistent with the choice of the E-scheme (Section 2.3), the angular distance $\Delta\mathcal{R} = \sqrt{\Delta Y^2 + \Delta\varphi^2}$, where Y is the true rapidity, is used.

A cone jet algorithm is infrared and collinear unsafe if it searches stable cones starting from seeds only, as was done in Run I. This can be cured [6] by using also "midpoints", *i.e.* combinations of at least two seeds, as starting positions in the proto-jet finding step (Section 3.2).

Finally, a merging/splitting procedure (Section 3.3) is applied to proto-jets sharing items in order to form the final jets, for which a minimal transverse momentum $p_{\mathrm{T}}^{\mathrm{min}}$ is required. The cone algorithm described schematically above and in more detail hereafter is called[4] the Midpoint algorithm in CDF and the Run II Cone algorithm in DØ.

3.1. Preclustering

The preclustering used in DØ at Run I was associating contiguous cells, a procedure that cannot be easily transposed to parton or particle jets. In Run II, the DØ preclustering groups items with $p_{\mathrm{T}} > 1$ MeV which do not already belong to a precluster in cones of radius $R = 0.3$ around items with $p_{\mathrm{T}} > 500$ MeV, treated in decreasing p_{T} order. It should be noted that no stability condition is required, *i.e.* the axis of the cone is not necessarily aligned with the direction of the precluster. This preclustering also helps to reduce the dependency of the jet algorithm on the detailed geometry of the DØ calorimeter, in particular the potential losses of efficiency due to the sharing of energy between adjacent towers.

3.2. Proto-jet finding

Items (CDF) or preclusters (DØ) of p_{T} above a given threshold (1 GeV) are considered as seeds. Around each seed[5], a proto-jet candidate is built from items in a cone of a given radius R_{cone} in $\Delta\mathcal{R}$. If the proto-jet and cone directions coincide, the proto-jet is added to the list of found proto-jets. Otherwise, the process is repeated[6] using the direction of the proto-jet as a new cone axis until the cone is stable.

During this search for a stable cone, items can flow in and out of the proto-jet, *e.g.* the seed (CDF) or items belonging to it (DØ) may not be part of the proto-jet found from this seed. Consequently, a group of nearby items of moderate total p_{T}, which would form a valid proto-jet if it was isolated, may be, at the same time, too close to and too far away from a higher p_{T} proto-jet to be part of any stable cone. This "lost jet" problem[7]

[4] It was called the "Improved Legacy Cone Algorithm" (ILCA) in [4].

[5] In DØ, preclusters are treated in decreasing p_{T} order and those which are closer than $R_{\mathrm{cone}}/2$ from an already found proto-jet are not considered as seeds.

[6] In DØ, this process stops if the p_{T} of the proto-jet candidate is lower than $p_{\mathrm{T}}^{\mathrm{min}}/2$.

[7] To solve this problem, the CDF Run I cone algorithm used "ratcheting", *i.e.* forcing each tower which had entered a proto-jet during the iterative proto-jet finding procedure to be kept in that proto-jet until the end of the procedure.

has been observed by CDF since Run I. In Run II, CDF uses the "Smaller Search Cone" solution [7], *i.e.* searching for stable cones with a smaller radius than that of the final proto-jets. However, proto-jets then do not generally correspond to stable cones anymore. This option is thus not used in DØ, where no lost jet has been observed yet.

Since a cone algorithm using seeds is sensitive to soft radiation, a similar[8] proto-jet finding procedure is applied using also midpoints as trial positions. To limit the number of midpoints, only those based on pairs of proto-jets satisfying the condition $\Delta\mathcal{R} < 2R_{\text{cone}}$ are used[9]. Using proto-jets instead of seeds as well as considering only pairs and not higher multiplicity combinations should make no difference in the result [4]. However, preventing midpoints to be formed when proto-jets are distant by more than $2R_{\text{cone}}$ is an extension of a cut which is only valid for seeds at the parton level. This may thus induce a (hopefully small) infrared unsafety at the detector level.

3.3. Merging/splitting

To avoid double counting of energy, all the proto-jets, formed from seeds as well as from midpoints, are considered[10] together in order to solve possible cases of items belonging to more than one proto-jet. When two proto-jets share items, they are combined into a single proto-jet (merging) or each shared item is exclusively attributed to its closest proto-jet (splitting). Each new proto-jet is then recalculated from its updated list of items. Merging (splitting) happens if the fraction f of the transverse momentum of the lower p_{T} proto-jet which is shared with the other proto-jet is larger (smaller) than a given parameter f_{min}. A value of $f_{\text{min}} = 50\%$ is used by both experiments[11]. Since the result may depend on the order of treatment of the proto-jets, *e.g.* in the case of three overlapping proto-jets [4], these are ordered by decreasing p_{T} before the process starts and re-ordered after each merging/splitting step. Jets with $p_{\text{T}} > p_{\text{T}}^{\text{min}}$ obtained at the end of this merging/splitting procedure are the objects used in physics analyses.

4. The $k_\perp$ algorithm

$k_\perp$ algorithms, whose first version [8] was proposed for e^+e^- physics, recursively cluster sets of items into larger sets: the initial sets consist of just one item each and the final sets are the jets. Besides being infrared

[8] In DØ, in contrast to preclusters, midpoints which are closer than $R_{\text{cone}}/2$ from an already found proto-jet are also considered as starting positions.

[9] In DØ, the condition $R_{\text{cone}} < \Delta\mathcal{R}$ is also required.

[10] In order to prevent the formation of large E_{T} jets from the combination of many unphysical low E_{T} proto-jets, a minimal E_{T} was required in Run I for proto-jets to enter the merging/splitting procedure. In contrast, no p_{T} cut is applied in Run II.

[11] In Run I, CDF was using a value of 75 %.

and collinear safe by design, $k_\perp$ algorithms ensure the factorisation of initial state radiation into universal (process independent) structure functions of the hadrons [9]. They also naturally avoid undesireable features of cone algorithms, such as unclustered energy or overlapping jets.

In contrast to the cone jets which are not modified during the merging/splitting process, the size of $k_\perp$ jets is not fixed in advance. This can lead to experimental difficulties in high luminosity environment or in case of calorimeter noise problems, since unphysical $k_\perp$ jets can then be formed from the clustering of many low energy clusters. However, this varying size may be an advantage for the energy calibration of $k_\perp$ jets, compared to cone jets, due to the absence of out-of-cone showering corrections [4].

The computation time of $k_\perp$ algorithms grows like N^3, where N is the number of items. In DØ, where there are typically several hundreds of towers with significant energy per event, a preclustering, which consists of simply grouping the towers in 2×2 fixed squares, is thus needed to reduce N.

The longitudinally invariant inclusive $k_\perp$ algorithm [10] is the most suited to inclusive measurements at hadron colliders and is thus recommended for Run II [4] in a slightly modified[12] version. It uses a distance parameter D, analogous to the cone radius R_{cone} for cone algorithms. Starting from a list of items (towers in CDF or preclusters in DØ for data), a distance to the beam axis $d_i = p_{\text{T}}^i$ is calculated for each item i and a distance $d_{ij} = \text{Min}(p_{\text{T}}^i, p_{\text{T}}^j) \times \Delta\mathcal{R}/D$ is calculated for each pair of items (i, j). If the smallest of all d_i and d_{ij} is a d_i, i is placed in the list of proto-jets and removed from the list of items. If it is a d_{ij}, i and j are combined according to the E-scheme (Section 2.3) into a new object k, which replaces i and j in the list of items. The process is then iterated with the new list of items until the list is exhausted. The proto-jets with p_{T} above a given threshold $p_{\text{T}}^{\text{min}}$ are the jets which give a picture of the hard scattering.

5. Summary and outlook

The two jet finding algorithms recommended for the analysis of Run II data and used in the CDF and DØ experiments, the Midpoint (CDF) or Run II Cone (DØ) algorithms and the $k_\perp$ algorithm have been presented.

Despite the efforts to define a common cone algorithm [4] in order to allow meaningful comparisons between the results of both experiments to be made, the Midpoint and Run II Cone algorithms are slightly different. The Midpoint algorithm uses the Smaller Search Cone option. The Run II Cone algorithm does not and presents additional features (Section 3.2) compared

[12] The recombination is based on the E-scheme instead of the Snowmass scheme and the distance between items uses the true rapidity Y instead of the pseudo-rapidity η.

to the recommendations of [4]. That these differences induce small effects on experimental results needs to be checked.

Even if its definition seems less intuitive at first sight, the $k_\perp$ jet algorithm is more straightforward than its cone equivalent. It is favoured theoretically, due to its built-in infrared and collinear safety and to its factorisation properties. It avoids problems associated to cone jet definition. The varying size of $k_\perp$ jets may also be an advantage for energy calibration. On the other hand, there might be more of unphysical jets reconstructed by the $k_\perp$ algorithm than by the cone algorithm in difficult experimental conditions, but the distance parameter D can certainly be adjusted to minimize this effect. Dedicated studies on this subject are needed: given the advantages of the $k_\perp$ algorithm, the development of its use at Tevatron is certainly worth the effort.

It is a pleasure to thank B. Gary and the organising committee of ISMD04 for a nice and fruitful meeting. I would also like to thank E. Busato, L. Duflot, B. Hirosky, M. Martinez-Perez, S. Tourneur and M. Wobisch for their help in preparing the talk and writing this report.

REFERENCES

[1] G. Sterman, S. Weinberg, *Phys. Rev. Lett.* **39**, 1436 (1977).

[2] F. Abe *et al.* [CDF Collaboration], *Nucl. Instrum. Methods* **A271**, 387 (1988); R. Blair *et al.* [CDF-II Collaboration], FERMILAB-PUB-96-390-E.

[3] S. Abachi *et al.* [D0 Collaboration], *Nucl. Instrum. Methods* **A338**, 185 (1994).

[4] G.C. Blazey *et al.*, `hep-ex/0005012`.

[5] J.E. Huth *et al.*, FERMILAB-CONF-90-249-E.

[6] R. Akers *et al.* [OPAL Collaboration], *Z. Phys.* **C63**, 197 (1994).

[7] S.D. Ellis, J. Huston, M. Tönnesmann, *eConf* **C010630**, P513 (2001) [`hep-ph/0111434`].

[8] S. Catani, Y.L. Dokshitzer, M. Olsson, G. Turnock, B.R. Webber, *Phys. Lett.* **B269**, 432 (1991).

[9] S. Catani, Y.L. Dokshitzer, B.R. Webber, *Phys. Lett.* **B285**, 291 (1992); S. Catani, Y.L. Dokshitzer, M.H. Seymour, B.R. Webber, *Nucl. Phys.* **B406**, 187 (1993).

[10] S.D. Ellis, D.E. Soper, *Phys. Rev.* **D48**, 3160 (1993) [`hep-ph/9305266`].