Available online at www.sciencedirect.com
ScienceDirect
Progress in Particle and Nuclear Physics 60 (2008) 484–551
Progress in Particle and Nuclear Physics
www.elsevier.com/locate/ppnp

Review

Jets in hadron–hadron collisions

S.D. Ellis[a,*], J. Huston[b], K. Hatakeyama[c], P. Loch[d], M. Tönnesmann[e]

[a] *University of Washington, Seattle, WA 98195, United States*
[b] *Michigan State University, East Lansing, MI 48824, United States*
[c] *Rockefeller University, New York, NY 10021, United States*
[d] *University of Arizona, Tucson, AZ 85721, United States*
[e] *Max Planck Institute fur Physics, Munich, Germany*

Abstract

In this article, we review some of the complexities of jet algorithms and of the resultant comparisons of data to theory. We review the extensive experience with jet measurements at the Tevatron, the extrapolation of this acquired wisdom to the LHC and the differences between the Tevatron and LHC environments. We also describe a framework (SpartyJet) for the convenient comparison of results using different jet algorithms.

Keywords: Jet; Jet algorithm; LHC; Tevatron; Perturbative QCD; SpartyJet

Contents

1. Introduction 485
2. Factorization 486
3. Jets: Parton level vs experiment 490
   3.1. Iterative cone algorithm 490
      3.1.1. Definitions 490
      3.1.2. R_{sep}, seeds and IR-sensitivity 495
      3.1.3. Seedless and midpoint algorithms 498
      3.1.4. Merging 499
      3.1.5. Summary 499

* Corresponding author.
E-mail address: ellis@phys.washington.edu (S.D. Ellis).

0146-6410/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ppnp.2007.12.002

3.2. k_T algorithm 500
3.3. Jet masses for jets at NLO 501
3.4. Recent cone algorithm issues 505
3.4.1. Jets at the "smeared" parton level and dark towers 505
3.4.2. The search cone algorithm 507
3.4.3. The midpoint cone algorithm with a "second pass" 508
3.4.4. Summary 508
3.5. Jets at the hadron level 509
4. Jets at the Tevatron 510
4.1. Detectors 511
4.2. Jet reconstruction and energy measurement 512
4.2.1. Jet algorithms 513
4.2.2. Jet energy scale at DØ 513
4.2.3. Jet energy scale at CDF 514
4.2.4. Summary 516
4.3. Monte Carlo tuning 516
4.3.1. Di-jet angular decorrelations 517
4.3.2. Underlying event 518
4.3.3. Jet shapes 519
4.3.4. Summary 520
4.4. Inclusive jet cross-sections 520
4.5. W/Z+jets 525
4.6. Heavy flavor jets 528
4.7. Particle flow type approaches 531
5. Jets at the LHC 531
5.1. Expectations for jet final states 532
5.2. Jet physics environment at the LHC 534
5.2.1. The underlying event at the LHC 534
5.2.2. Minimum-bias events and pile-up 535
5.3. Experimental aspects of jet reconstruction at the LHC 536
5.3.1. Brief look at the ATLAS and CMS calorimeters 536
5.3.2. Calorimeter jet basics 538
5.3.3. Calorimeter jet calibration 539
5.3.4. Use of tracks in jets 541
5.3.5. Jet algorithms 541
5.4. Jet signal characteristics at the LHC 541
6. SpartyJet 545
7. Conclusions 548
Acknowledgements 549
References 549

1. Introduction

Most of the interesting physics signatures at the Tevatron and LHC involve final states with jets of hadrons. A jet is reconstructed from energy depositions in calorimeter cells and/or from charged particle track momenta, and ideally is corrected for detector response and resolution effects so that the resultant 4-vector corresponds to that of the sum of the original hadrons comprising the jet. The jets can also be further corrected, for hadronization effects, back to the parton(s) from which the jet originated. The resultant measurements can be compared to

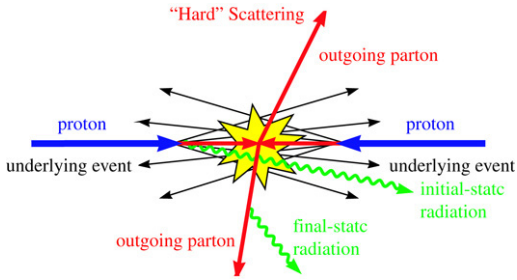


Fig. 1. Pictorial representation of a hard scattering event [2].

predictions from parton-shower Monte Carlos. If further corrections are made to account for the showers, or if these corrections are assumed to be small, comparisons can be made directly to the short-distance partons described by fixed-order perturbative calculations.

In order to actually reconstruct a jet, and make comparisons between data and theoretical predictions, a precise definition of the jet is required. The definition is presented in the form of a jet algorithm. Jet algorithms cluster partons, or particles or calorimeter towers based on proximity in coordinate space (as for example in cone algorithms) or proximity in momentum space (as for example in k_T algorithms). For a precise comparison of experiment to theory, it is advantageous for a jet algorithm to provide a similar description of a hard scatter event regardless if it is applied at the detector, hadron or parton level.

In this article, we will review some of the complexities of jet algorithms and of the resultant comparisons of data to theory. We will review the extensive experience with jet measurements at the Tevatron, the extrapolation of this acquired wisdom to the LHC and the differences between the Tevatron and LHC environments. We will also discuss ways in which the jet algorithm systematics can be reduced to the percent level, an important goal for LHC analyses. Finally we will describe a framework (SpartyJet) for the convenient comparison of results using different jet algorithms. Several of the authors are members of CDF and ATLAS and we apologize in advance for our concentration on those experiments. Given the restrictions of space, we will not try a comprehensive review of physics with jets at hadron–hadron colliders, but instead refer the reader to a recent review [1].

2. Factorization

The fundamental challenge when trying to make theoretical predictions or interpret experimentally observed final states is that the theory of the strong interactions (QCD) is most easily applied to the short-distance ($\ll$1 fermi) degrees of freedom, *i.e.*, to the color-charged quarks and gluons, while the long-distance degrees of freedom seen in the detectors are color singlet bound states of these degrees of freedom. We picture the overall scattering process, as pictorially displayed in Fig. 1 [2], evolving from the incoming long-distance hadrons in the beams to the short-distance scattering process to the long-distance outgoing states, as occurring in several (approximately) distinct steps. The separation, or factorization, of these steps is essential both conceptually and calculationally. It is based on the distinct distance (or momentum) scales inherent at each step.

We imagine as a first step picking out from the incident beam particles the short-distance partons (defined by an appropriate factorization scale) that participate in the short-distance scattering. The relative probability to find the scattering partons at this step is provided by the parton distribution functions (pdfs), which are functions of the partons' color and flavor, the

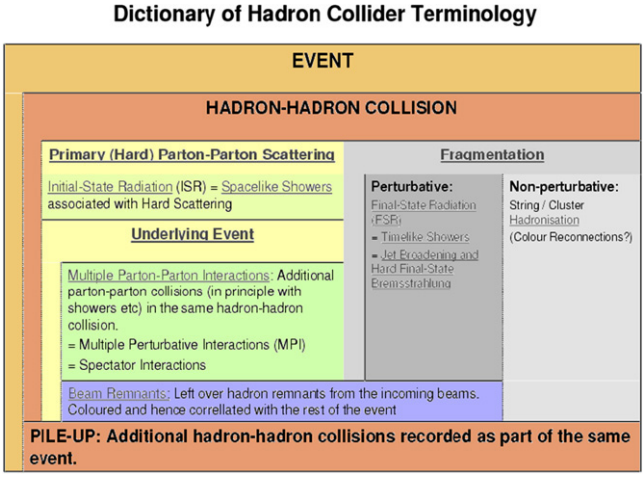


Fig. 2. A dictionary of hadron collider terms relating to jet measurements [6].

longitudinal momentum fractions x_k carried by the partons and the factorization scale μ, all of which serve to uniquely define the desired partons. The pdfs are themselves determined from global fits [3] to a wide variety of data, all of which can be analyzed in the context of perturbative QCD (pQCD) essentially as outlined here. The partons selected in this way can emit radiation prior to the short-distance scattering yielding the possibility of initial-state radiation (ISR). The remnants of the original hadrons, with one parton removed, are no longer color singlet states and will interact, presumably softly, with each other generating (approximately incoherently from the hard scattering) an underlying distribution of soft partons, the beginning of the underlying event (UE).

Next comes the short distance, large momentum transfer scattering process that may change the character of the scattering partons and/or produce more partons (or other interesting particles). The cross-section for this step is evaluated at fixed order in pQCD, presumably to next-to-leading order (NLO), or higher. Then comes another color radiation step, when many new gluons and quark pairs are added to the state (final-state radiation or FSR), dominated by partons that have low energy and/or are nearly collinear with the scattered short-distance partons. The FSR, like the ISR, is described by the showers in the Monte Carlo programs [4,5] and calculated probabilistically in terms of summed leading logarithm perturbation theory.

The final step in the evolution to long-distance states involves a nonperturbative hadronization process that organizes the colored degrees of freedom from the showering and from the softer interactions of other initial-state partons (the UE is simulated in terms of beam–beam remnants and multiple parton interactions) into color-singlet hadrons with physical masses. This nonperturbative hadronization step is accomplished in a model dependent fashion (*i.e.*, different fashion) in different Monte Carlos. The resulting hadrons comprise a collection of both ground-state hadrons (primarily pions) and resonances (A_1, A_2, etc), and the resonances then decay into lighter hadrons such as pions. The masses of the resonances result in the decay pions being produced with nonzero transverse momenta with respect to the momentum direction of the original resonance (but still small compared to typical jet momenta). The details are determined from data but the decay modes for some of the higher mass resonances are not well-understood leading to uncertainties (and differences between Monte Carlo programs) in the details of the

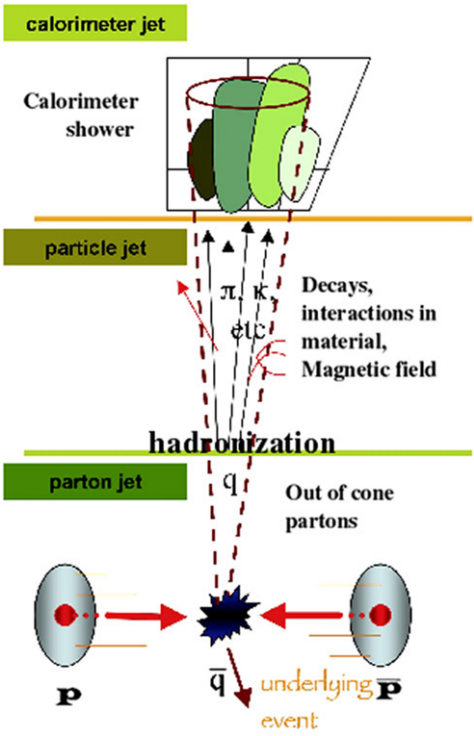


Fig. 3. A representation of the stages of jet production and measurement.

production of the final-state hadrons. A convenient dictionary of the terms described above is summarized in Fig. 2 [6].

The union of the showering and the hadronization steps is what has historically been labeled as fragmentation, as in fragmentation functions describing the longitudinal distribution of hadrons within final-state jets. In practice, both the radiation and hadronization steps tend to smear out the energy that was originally localized in the individual short-distance partons (a "splash-out" effect), while the contributions from the underlying event (and any "pile-up" from multiple hadron collisions in the same data-taking time interval) add to the energy originally present in the short-distance scattering (a "splash-in" effect). Finally the hadrons, and their decay products, are detected with finite precision in a detector. This final step of the jet components interacting in the detector is represented in Fig. 3.

The separation of this complicated scattering process into distinct steps is not strictly valid in quantum mechanics where interference plays a role; we must sum the amplitudes before squaring them and not just sum the squares. However, some features of this factorization can be rigorously established [7], and the numerical dominance of collinear QCD radiation ensures that the simple picture presented here and quantified by Monte Carlo generated events, with interference only approximately represented in the structure of the showers, *e.g.*, the angles of emission of partons in the showers are monotonically ordered, provides a reliable first approximation.

In order to interpret the experimentally detected long-distance objects, the charged particles and energy depositions in calorimeter cells, in terms of the underlying short-distance physics, the partons, jet algorithms are employed to associate "nearby" experimental objects into jets. The

underlying assumption is that the kinematics (energy and momentum) of the resulting cluster or jet provides a useful measure of the kinematics (energy and momentum) of the underlying, short-distance partons. The goal is to characterize the short-distance physics, event-by-event, in terms of the discrete jets found by the algorithm. In particular, we assume that the basic mismatch between colored short-distance theory objects and the colorless long-distance experimental objects does not present an important numerical limitation, compared to our goal of percent level accuracy for jet algorithm effects. We assume that we are either insensitive to or can reliably correct for the effects of the UE, ISR, FSR and hadronization. We assume that most relevant features of a jet can be described by the (up to) 2 partons per jet present in a NLO calculation. We will investigate the features that are not so well-described.

As noted, jet algorithms rely on the merging of objects that are, by some measure, nearby each other. This feature is essential in perturbation theory, where the divergent contributions from virtual diagrams must contribute in exactly the same way, *i.e.*, contribute to the same kinematic bins, as the divergent contributions from soft and collinear real emissions, in order that these contributions can cancel. It is only through this cancellation that jet algorithms serve to define an IR-safe (finite) quantity, *i.e.*, a quantity that is insensitive to the emission of extra soft and/or collinear partons. The standard measures of "nearness" (see [8]) include pair-wise relative transverse momenta, as in the k_T algorithm, or angles relative to a jet axis, as in the cone algorithm. By definition a "good" algorithm yields stable (*i.e.*, very similar) results whether it is applied to a state with just a few partons (as in NLO perturbation theory), a state with many partons (after the parton shower as simulated in a Monte Carlo), a state with hadrons (as simulated in a Monte Carlo including a model for the hadronization step and the underlying event), or applied to the observed tracks and energy deposition in a real detector. As we will see, this requirement constitutes a substantial challenge. Further, it is highly desirable that the identification of jets be insensitive to the contributions from the simultaneous uncorrelated soft collisions that occur during pile-up at high luminosity. Finally, we want to be able to apply the *same* algorithm (in detail) at each level in the evolution of the hadronic final state. This implies that we must avoid components in the algorithm that make sense when applied to data but not to fixed-order perturbation theory, or vice versa. This constraint will play an important role in our subsequent discussion.

In practice, we can think of the jet algorithm as a set of mathematical rules that detail how to carry out two distinct steps. The first step operates, event-by-event, on the list of 4-vectors, which describes either the perturbative final state, the final-state hadrons in the MC simulated event or the output from the detector, to turn the original list into a set of sublists, one sublist for each jet (plus the beam jets). The second step specified by the algorithm tells us how to construct appropriate kinematic quantities from each sublist in order to describe the kinematic properties of the individual jets. Both steps depend on the specific jet algorithm, as will be illustrated in detail in the next sections for specific jet algorithms, and have varied over time. In particular, early applications of jets did not employ true 4-vector arithmetic and largely ignored the information carried by the invariant mass of the jet. In Run II at the Tevatron and at the LHC true 4-vector arithmetic is and will be employed (the so-called "E-scheme" as recommended for Run II in [8]). The corresponding kinematic variables describing the jets include the (true) transverse momentum, p_T, the rapidity, $y = \frac{1}{2}\log\frac{E+p_Z}{E-p_Z}$, the azimuthal angle, ϕ, and the invariant mass of the jet, M_J. We will have more to say about the history of different choices of kinematic variables in the next section.

For many events, the jet structure is clear and the jets, into which the individual calorimeter towers should be assigned, are fairly unambiguous. However, in other events, such as the lego

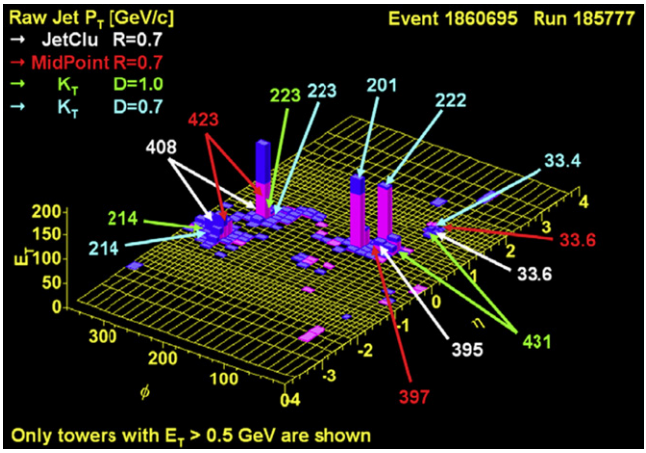


Fig. 4. Impact of different jet clustering algorithms on an interesting CDF event taken in Run II. The segmentation in η and ϕ shown in the lego plot corresponds to the calorimeter segmentation. Energy depositions in the electromagnetic portion of the calorimeter are colored red and those in the hadronic section are colored blue. The numbers in the figure are transverse momenta of the jets pointed to by the arrows, and different colors represent jets clustered by different algorithms.

plot of a CDF event shown in Fig. 4, the complexity of the energy depositions means that different algorithms will result in different assignments of towers to the various jets. This is not a problem if a similar complexity is exhibited by the theoretical calculation which is to be compared to the data. However, the most precise and thoroughly understood theoretical calculations arise in fixed-order perturbation theory, which can exhibit only limited complexity, *e.g.*, at most 2 partons per jet at NLO. On the other hand, for events simulated with parton-shower Monte Carlos the complexity of the final state is more realistic, but the intrinsic theoretical uncertainty is larger. Correspondingly the jets identified by the algorithms vary if we compare at the perturbative, shower, hadron and detector levels. Thus it is essential to understand these limitations of jet algorithms and, as much as possible, eliminate or correct for them to approach our percent level goal. It is the aim of the following review to highlight the issues that arose during Runs I and II at the Tevatron, discuss their current understanding, and outline possible preventative measures for the LHC [6].

3. Jets: Parton level vs experiment

3.1. Iterative cone algorithm

3.1.1. Definitions

To illustrate the behavior of jet algorithms consider first the original Snowmass implementation of the iterative cone algorithm [9]. The first step in the algorithm, *i.e.*, the identification of the sublists of objects corresponding to the jets, is defined in terms of a simple sum over all (short-distance or long-distance) objects within a cone centered at rapidity (the original version used the pseudorapidity $\eta = \frac{1}{2}\log(\cot\frac{\theta}{2})$) and azimuthal angle (y_C, ϕ_C). Using the objects in the cone we can define a p_T-weighted centroid via

$$k \subset C \quad \text{iff} \quad \sqrt{(y_k - y_C)^2 + (\phi_k - \phi_C)^2} \leq R_{\text{cone}},$$

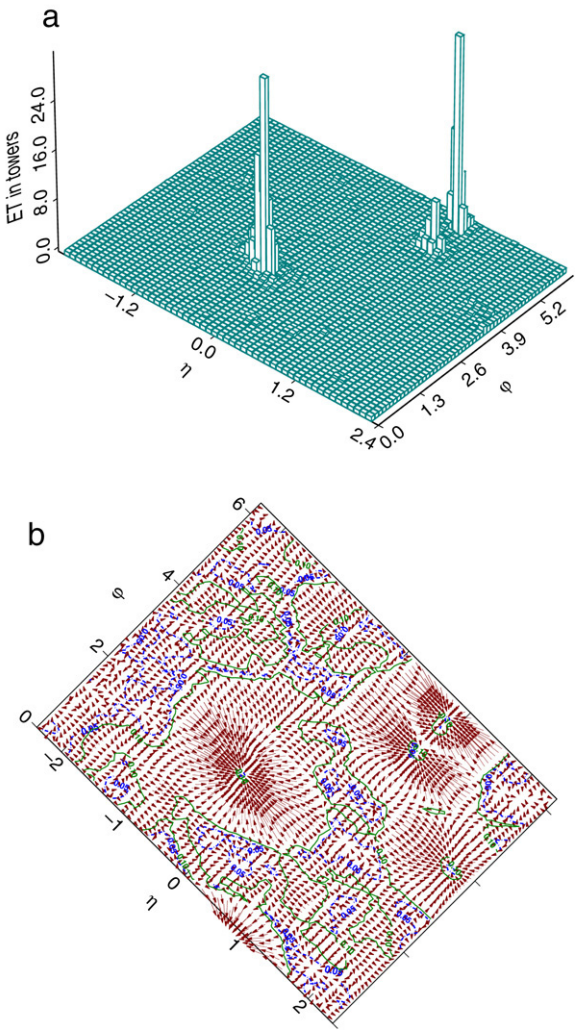


Fig. 5. (An ideal) Monte Carlo generated event with 2 large energy jets and 1 small energy jet in the LEGO plot (a), and the corresponding flow structure of the trial cones in (b).

$$\overline{y}_C \equiv \frac{\sum_{k\subset C} y_k \cdot p_{T,k}}{\sum_{l\subset C} p_{T,l}}, \qquad \overline{\phi}_C \equiv \frac{\sum_{k\subset C} \phi_k \cdot p_{T,k}}{\sum_{l\subset C} p_{T,l}}.$$

If the p_T-weighted centroid does not coincide with the geometric center of the cone, $(\overline{y}_C, \overline{\phi}_C) \neq (y_C, \phi_C)$, a cone is placed at the p_T-weighted centroid and the calculation is repeated. This simple calculation is *iterated* until a "stable" cone is found, $(\overline{y}_C, \overline{\phi}_C) = (y_C, \phi_C)$, which serves to define the jet (and the name of this algorithm as the iterative cone algorithm). Thus, at least in principle, one can think in terms of placing trial cones everywhere in (y, ϕ) and allowing them to "flow" until a stable cone or jet is found. This flow idea is illustrated in Fig. 5, where (a) illustrates the LEGO plot for a simple (quiet) Monte Carlo generated event with 3 apparent

jets and (b) shows the corresponding flows of the trial cones towards the (obvious) final jets. Compared to the event in Fig. 4 there is little ambiguity in this event concerning the jet structure.

To facilitate the subsequent discussion and provide some mathematical structure for this image of "flow" we can define the "Snowmass potential" in terms of the 2-dimensional vector $\overrightarrow{r} = (y, \phi)$ via

$$V\left(\overrightarrow{r}\right) = -\frac{1}{2}\sum_k p_{T,k}\left(R^2_{\rm cone} - \left(\overrightarrow{r}_k - \overrightarrow{r}\right)^2\right)\Theta\left(R^2_{\rm cone} - \left(\overrightarrow{r}_k - \overrightarrow{r}\right)^2\right),$$

where the Θ function defines the objects inside the cone centered at r. The flow described by the iteration process is driven by the "force"

$$\begin{aligned}\overrightarrow{F}\left(\overrightarrow{r}\right) &= -\overrightarrow{\nabla} V\left(\overrightarrow{r}\right) = \sum_k p_{T,k}\left(\overrightarrow{r}_k - \overrightarrow{r}\right)\Theta\left(R^2_{\rm cone} - \left(\overrightarrow{r}_k - \overrightarrow{r}\right)^2\right)\\ &= \left(\overrightarrow{\overline{r}}_{C\left(\overrightarrow{r}\right)} - \overrightarrow{r}\right)\sum_{k\subset C(r)} p_{T,k},\end{aligned}$$

where $\overrightarrow{\overline{r}}_{C(\overrightarrow{r})} = (\overline{y}_{C(\overrightarrow{r})}, \overline{\phi}_{C(\overrightarrow{r})})$ and $k \subset C(\overrightarrow{r})$ is defined by $\sqrt{(y_k - y)^2 + (\phi_k - \phi)^2} \leq R_{\rm cone}$. As desired, this "force" pushes the cone to the stable cone position, *i.e.*, the minimum of the Snowmass potential. As noted above the current Run II analyses and those expected at the LHC, described in more detail below, 4-vector techniques are used and the corresponding E-scheme centroid is given instead by

$$\begin{aligned}&k \subset C \quad \text{iff } \sqrt{(y_k - y_C)^2 + (\phi_k - \phi_C)^2} \leq R_{\rm cone},\\ &p_C = \left(E_C, \overrightarrow{p}_C\right) = \sum_{k\subset C}\left(E_k, \overrightarrow{p_k}\right), \qquad \overline{y}_C \equiv \frac{1}{2}\ln\frac{E_C + p_{z,C}}{E_C - p_{z,C}}, \qquad \overline{\phi}_C \equiv \tan^{-1}\frac{p_{y,C}}{p_{x,C}}.\end{aligned}$$

In the NLO perturbative calculation these changes in definitions result in only tiny numerical changes, compared to the Snowmass definition numbers.

As an introduction to how the iterative cone algorithm works, consider first its application to NLO level in perturbation theory (see, *e.g.*, [10]), where there are at most 2 partons in a cone. As defined above, the cone algorithm specifies that two partons are included in the same jet (*i.e.*, form a stable cone) if they are both within $R_{\rm cone}$ (*e.g.*, 0.7 in (y, ϕ) space) of the centroid, which they themselves define. This means that 2 partons of equal p_T can form a single jet as long as their pair-wise angular separation does not exceed the boundaries of the cone, $\Delta R = 2R_{\rm cone}$. On the other hand, as long as $\Delta R > R_{\rm cone}$, there will also be stable cones centered around each of the partons. The corresponding 2-parton phase space for $R_{\rm cone} = 0.7$ is illustrated in Fig. 6(a) in terms of the ratio $z = p_{T,2}/p_{T,1}\left(p_{T,1} \geq p_{T,2}\right)$ and the angular separation variable $d = \sqrt{(y_1 - y_2)^2 + (\phi_1 - \phi_2)^2}$. To the left of the line $d = R_{\rm cone}$ the two partons always form a single stable cone and jet, while to the far right, $d > 2R_{\rm cone}$, there are always two distinct stable cones and jets, with a single parton in each jet. More interesting is the central region, $R_{\rm cone} < d < 2R_{\rm cone}$, which exhibits both the case of two stable cones (one of the partons in each cone) and the case of three stable cones (the previous two cones plus a third cone that includes both partons). The precise outcome depends on the specific values of z and d. (Note that the exactly straight diagonal boundary in the figure corresponds to the p_T-weighted definition of the Snowmass algorithm, but is only slightly distorted, <2%, when full 4-vector kinematics is used in the Run II algorithms.) To see the three stable cone structure in terms of the 2-parton

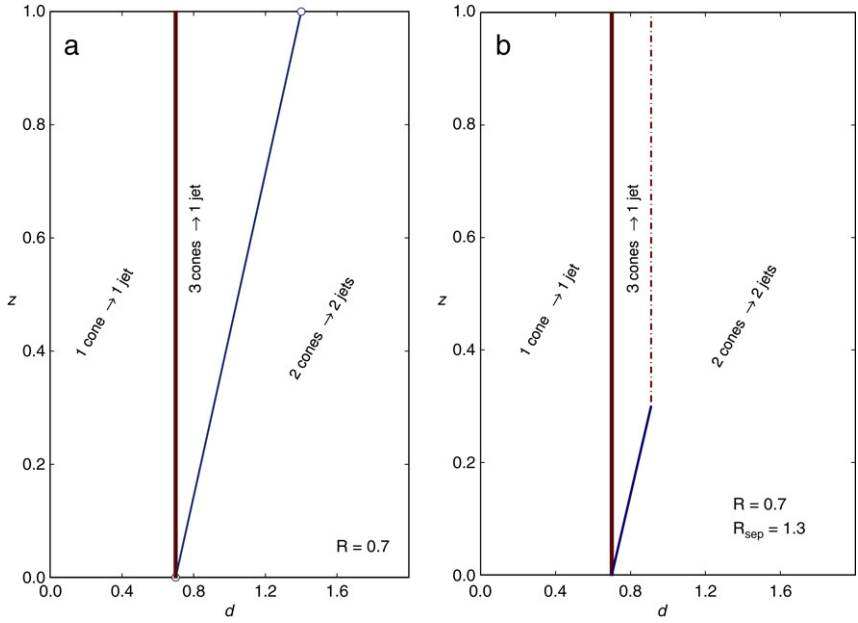


Fig. 6. Perturbative 2-parton phase space: $z = p_{T,2}/p_{T,1}\,(p_{T,1} \geq p_{T,2})$, $d = \sqrt{(y_1 - y_2)^2 + (\phi_1 - \phi_2)^2}$ for (a) the naive $R_{\text{sep}} = 2$ case and (b) for $R_{\text{sep}} = 1.3$ case suggested by data.

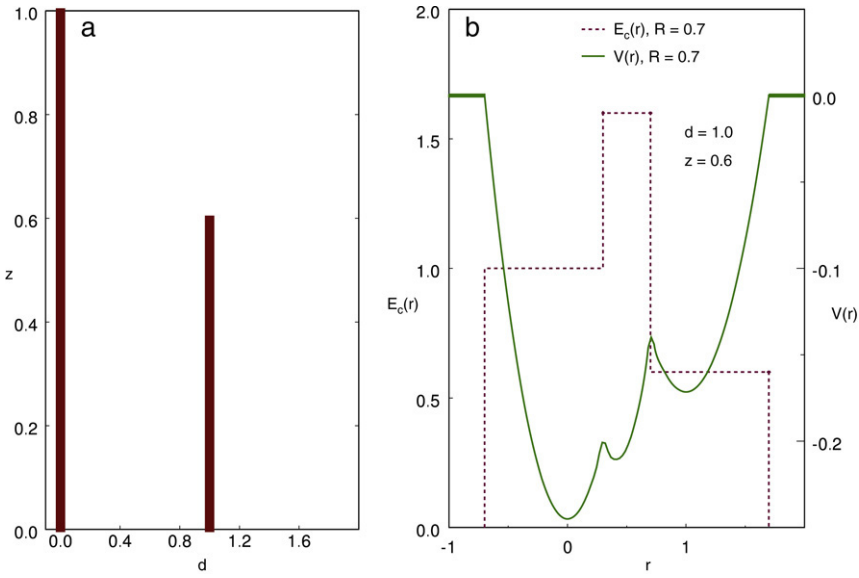


Fig. 7. 2-parton distribution in (d, z) in a) with $d = 1.0$, $z = 0.6$ and the corresponding energy-in-cone, $E_C\,(r)$, and potential, $V\,(r)$.

"Snowmass potential" consider the point $z = 0.6$ and $d = 1.0$, which is in the 3 cones → 1 jet region. This configuration, in terms of the 2-parton coordinates z and d, is illustrated in Fig. 7(a).

The corresponding energy in a cone, E_c, normalized to the energy of the more energetic parton, is illustrated by the (red) dashed line in Fig. 7(b), where the location of the cone center, r, is constrained to lie in the plane of the 2 partons with the origin at the location of the more energetic parton. The solid (green) curve in Fig. 7(b) shows the corresponding "Snowmass potential", again normalized to the energy of the more energetic parton, versus the same 1-dimensional location radius r. This potential exhibits the expected 3 minima corresponding to a stable cone at each parton ($r = 0$ and $r = d = 1.0$) and a third stable cone, central between the other two, that includes the contributions from both partons. A relevant point is that the central minimum is not nearly as deep (*i.e.*, as robust) as the other two. As we shall see, this minimum often does not survive the smearing inherent in the transition from the short distances of fixed-order perturbation theory to the long distances of the physical final state. As indicated by the labeling in Fig. 6, in the 3 stable cone region the original perturbative calculation [10] is kept as the jet in the 2-in-1 stable cone, maximum p_T configuration, *i.e.*, the cone that included all of the energy in the other two cones consistent with the merging discussion below.

As we will see, much of the concern and confusion about the cone algorithm arises from the treatment of this 3 stable cone region. It is intuitively obvious that, as the energy in the short-distance partons is smeared out by subsequent showering and hadronization, the detailed structure in this region is likely to change. In particular, while two nearly equal p_T partons separated by nearly $2R_{\rm cone}$ may define a stable cone in fixed-order pQCD, this configuration is unlikely to yield a stable cone after showering and hadronization.

Having performed the first step in the algorithm to identify the particles in the cone C, the second step is to define the kinematic variables describing the jet. As suggested above the Snowmass definition of the iterative cone algorithm used angular variables defined by the p_T-weighted expressions

$$\overline{y}_J \equiv \frac{\sum\limits_{k\subset C} y_k \cdot p_{T,k}}{\sum\limits_{l\subset C} p_{T,l}}, \qquad \overline{\phi}_J \equiv \frac{\sum\limits_{k\subset C} \phi_k \cdot p_{T,k}}{\sum\limits_{l\subset C} p_{T,l}},$$

and, instead of the true transverse momentum, the scalar transverse momentum

$$P_J = p_{T,\text{Snowmass}} = \sum_{l\subset C} p_{T,l}.$$

In the (recommended) case of 4-vector arithmetic we have instead (as suggested as above)

$$p_C = \left(E_C, \overrightarrow{p}_C\right) = \sum_{k\subset C} \left(E_k, \overrightarrow{p_k}\right), \qquad y_J \equiv \frac{1}{2}\ln\frac{E_C + p_{z,C}}{E_C - p_{z,C}}, \qquad \phi_J \equiv \tan^{-1}\frac{p_{y,C}}{p_{x,C}},$$

$$P_J = \left|\overrightarrow{p}_{C,T}\right| = \left|\sum_{k\subset C} \overrightarrow{p}_{k,T}\right|, \qquad M_J^2 = p_C^2 = E_C^2 - \overrightarrow{p}_C^{\,2}.$$

During Run I CDF used an intermediate set of definitions with E-scheme (4-vector) angles and the jet "transverse momentum" given by the transverse energy

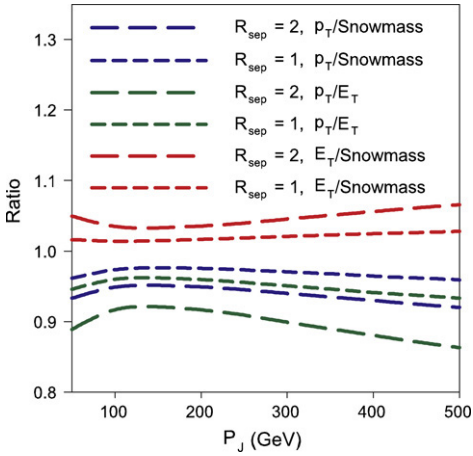
$$P_J = E_T = E_C \sin\theta_C.$$

In the next section we will briefly explore the quantitative differences between these definitions in the context of NLO perturbation theory.

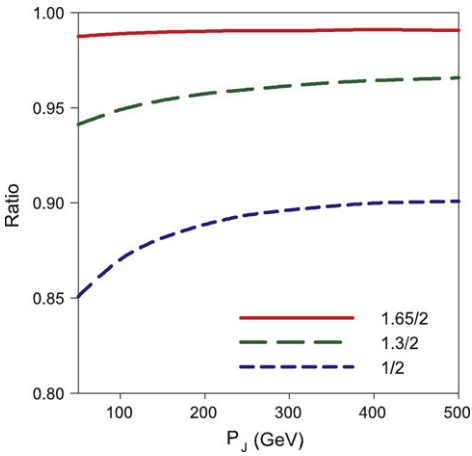
3.1.2. $R_{\rm sep}$, seeds and IR-sensitivity

Iterative cone algorithms similar to the ones described in the previous section were employed by both the CDF and DØ collaborations during Run I with considerable success. There was fairly good agreement with NLO pQCD for the inclusive jet cross-section over a dynamic range of order 10^8. During Run I, the data were corrected primarily for detector effects and for the contributions of the underlying event. In fact, a positive feature of the cone algorithm is that, since the cone's geometry in (y, ϕ) space is (meant to be) simple, the correction for the "splash-in" contribution of the (largely uncorrelated) underlying event (and pile-up) is straightforward. (As we will see below, the corrections being used in Run II are more sophisticated.) The uncertainties in both the data and the theory were 10% or greater, depending on the kinematic regime, and helped to ensure agreement. However, as cone jets were studied in more detail, various troubling issues arose. For example, it was noted long ago [11] that, when using the experimental cone algorithms implemented at the Tevatron, two jets of comparable energy (taken from 2 different events in the data) are not merged into a single jet if they are separated by an angular distance greater than approximately 1.3 times the cone radius, while the simple picture of Fig. 6(a) suggests that merging should occur out to an angular separation of $2R_{\rm cone}$. Independently it was also noted that the dependence of the experimental inclusive jet cross-section on the cone radius $R_{\rm cone}$ [12] and the shape of the energy distribution within a jet [13] both differed discernibly from the NLO predictions (the data were less $R_{\rm cone}$ dependent and exhibited less energy near the edge of the cone). All three of these issues seemed to be associated with the contribution from the perturbative configuration of two partons with comparable p_T at opposite sides of the cone ($z \simeq 1$, $d \simeq 2R_{\rm cone} = 1.4$ in Fig. 6(a)) and the data suggested a lower contribution from this configuration than present in the perturbative result. To simulate this feature in the perturbative analysis, a phenomenological parameter $R_{\rm sep}$ was added to the NLO implementation of the cone algorithm [14]. In this "experiment-aware" version of the perturbative cone algorithm, two partons are not merged into a single jet if they are separated by more than $R_{\rm sep} \cdot R_{\rm cone}$ from each other, independent of their individual distance from the p_T-weighted jet centroid. Thus, the two partons are merged into the same jet if they are within $R_{\rm cone}$ of the p_T-weighted jet centroid and within $R_{\rm sep} \cdot R_{\rm cone}$ of each other; otherwise the two partons are identified as separate jets. In order to describe the observed $R_{\rm cone}$ dependence of the cross-section and the observed energy profile of jets, the specific value $R_{\rm sep} = 1.3$ was chosen (along with a "smallish" renormalization/factorization scale $\mu = p_T/4$), which was subsequently noted to be in good agreement with the aforementioned (independent) jet separation study. The resulting 2-parton phase space is indicated in Fig. 6(b). In the perturbative calculation, this redefinition, leading to a slightly lower average p_T for the leading jet, lowers the NLO jet cross-section by about 5% (for $R = 0.7$ and $p_T = 100$ GeV/c). It is important to recognize that the fractional contribution to the inclusive jet cross-section of the merged 2-parton configurations in the entire wedge to the right of $d = R_{\rm cone}$ is only of order 10% for jet p_T of order 100 GeV/c, and, being proportional to $\alpha_s\,(p_T)$, decreases with increasing p_T. Thus it is no surprise that, although this region was apparently treated differently (in the NLO theory comparisons) by the cone algorithm implementations of CDF and DØ during Run I as discussed below, there were no relevant cross-section disagreements above the $>10\%$ uncertainties. Further, as we will discuss below, it is the variation in the treatment of this (effectively 10%) region of the 2-parton phase space that drives many of the differences between jet algorithms.

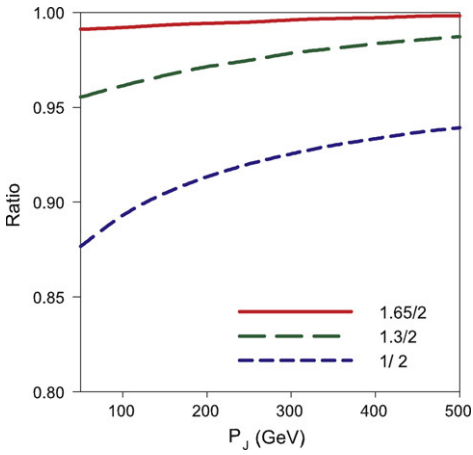
While the parameter $R_{\rm sep}$ is *ad hoc* and thus an undesirable complication in the perturbative jet analysis, it will serve as a useful pedagogical tool in the following discussions. To illustrate



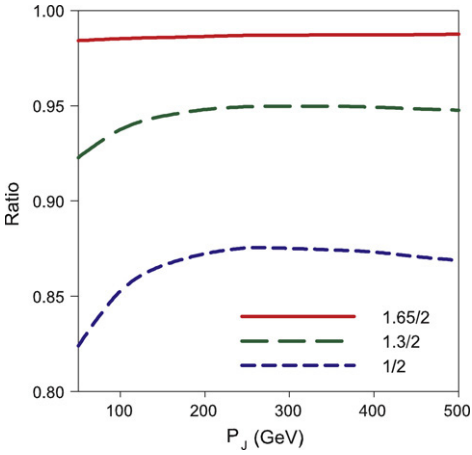
(a) Cross-section ratios.



(b) Snowmass definition.



(c) $p_T E$-scheme definition.



(d) $E_T E$-scheme definition.

Fig. 8. Ratios of the NLO inclusive cone jet cross-section versus the jet momentum for 3 definitions of the kinematics for various values of R_{sep}.

this point quantitatively, Fig. 8 shows the dependence on R_{sep} for various choices of the jet momentum P_J at NLO in perturbation theory. The curves labeled Snowmass use the p_T-weighted kinematics described above with P_J given by the scalar sum of the transverse momenta of the partons in the cone. The two E-scheme algorithms use full 4-vector kinematics and P_J equal to either the magnitude of the true (vector sum) transverse momentum (the recommended choice), or the "transverse energy" defined by $P_J = E_T = E \sin\theta$ (as defined by CDF in Run I). Thus this last variable knows about both the momentum and the invariant mass of the set of partons in the cone, which can be sizable for well-separated parton pairs. The differences in the various ratios for different values of R_{sep} tell us about how the 2-parton configurations contribute. For example, Fig. 8(a) tells us that, since, for a given configuration of 2 partons in a cone, $E_T > p_{T,\text{Snowmass}} > p_T$, the cross-sections at a given value of P_J will show the same ordering. Further, as expected, the differences are reduced if we keep only configurations

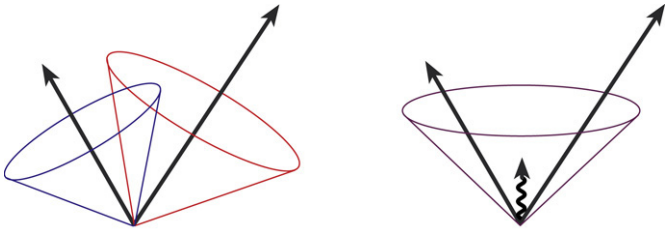


Fig. 9. Two partons in two cones or in one cone with a (soft) seed present [8].

with small angular separation, $R_{\rm sep} = 1$. From Fig. 8(b) we confirm the earlier statement that lowering $R_{\rm sep}$ from 2 to 1.3 yields a 5% change for the Snowmass algorithm cross-section with $P_J = 100$ GeV, while lowering it all the way to $R_{\rm sep} = 1$, *i.e.*, removing all of the triangular region, lowers the 100 GeV Snowmass jet cross-section by approximately 12%. Fig. 8(c) and (d) confirm that 4-vector kinematics with $P_J = p_T$ exhibits the smallest sensitivity to $R_{\rm sep}$, *i.e.*, to the 2-parton configurations in the triangle. The choice $P_J = E_T$, with its dependence on the mass of the pair, exhibits the largest sensitivity to $R_{\rm sep}$. These are all good reasons to use the recommended E-scheme kinematics with $P_J = p_T$. The move to employ 4-vector kinematics for jet analyses in Run II and at the LHC is a positive step. It will allow the meaningful investigation of jet masses, which will likely be very useful at the LHC. The decay of large mass (few TeV) new particles will lead to highly boosted W's, Z's and top quarks that will be observed as single jets. Thus the mass of such jets may constitute a useful selection tool.

The difference between the perturbative implementation of the iterative cone algorithm and the experimental implementation at the Tevatron, which is simulated by $R_{\rm sep}$, is thought to arise from several sources. While the perturbative version (with $R_{\rm sep} = 2$) analytically includes *all* 2-parton configurations that satisfy the algorithm (recall Fig. 6(a)), the experiments employ the concept of *seeds* to reduce the analysis time and place trial cones only in regions of the detector where there are seeds, *i.e.*, pre-clusters with substantial energy. This procedure introduces another parameter, the lower p_T threshold defining the seeds, and also decreases the likelihood of finding the 2-showers-in-one-jet configurations corresponding to the upper right-hand corner of the 3 cones $\rightarrow$ 1 jet region of Fig. 6(a) and the middle minimum in Fig. 7(b). Thus, the use of seeds contributes to the need for $R_{\rm sep} < 2$. Perhaps more importantly, the desire to match the theoretical algorithm with the experimental one means that we should include seeds in the perturbative algorithm. This is accomplished by placing trial cones only at the locations of each parton and testing to see if any other partons are inside these cones. Thus at NLO, 2 partons will be merged into a single jet only if they are closer than $R_{\rm cone}$ in (y, ϕ) space. This corresponds to $R_{\rm sep} = 1.0$ in the language of Fig. 6 and produces a larger numerical change in the analysis than observed, *i.e.*, we wanted $R_{\rm sep} \simeq 1.3$. More importantly at the next order in perturbation theory, NNLO, there are extra partons that can play the role of low energy seeds. The corresponding parton configurations are illustrated in Fig. 9. At NLO, or in the virtual correction in NNLO, the absence of any extra partons to serve as a seed leads to two distinct cones as on the left, while a (soft) real emission at NNLO can lead to the configuration on the right where the soft gluon "seeds" the middle cone that includes all of the partons. The resulting separation between the NNLO virtual contribution and the NNLO soft real emission contribution (*i.e.*, they contribute to different jet configurations) leads to an undesirable logarithmic dependence on the seed p_T threshold [15]. In the limit of an arbitrarily soft seed p_T cutoff, the cone algorithm with seeds is no longer IR-safe. By introducing seeds in the algorithm, we have introduced exactly what we want to avoid,

sensitivity to soft emissions. From the theory perspective, seeds are an undesirable component in the algorithm and should be eliminated.

3.1.3. Seedless and midpoint algorithms

The labeling of the Run I cone algorithm with seeds as Infrared Unsafe has led some theorists to suggest that the corresponding analyses should be disregarded. This is too extreme a reaction, especially since the numerical difference between the jet cross-section found in data using seeds is expected to be less than a few percent different from an analysis using a seedless algorithm. A more useful approach will be to either avoid the use of seeds,[1] or to correct for them in the analysis of the data, which can then be compared to a perturbative analysis without seeds. Note that it may seem surprising that an algorithm, which is Infrared Unsafe due to the use of seeds, leads to experimental results that differ from an Infrared-Safe seedless algorithm by only a few percent. The essential point is that the lack of IR-safety is a property of the fixed-order perturbative application of the algorithm with seeds, not of the experimental application. In real data the additional soft components of the event (initial-state radiation, final-state radiation and the underlying event) ensure that there are seeds "nearly" everywhere. Thus there is only a small change from the situation where seeds are assumed to be everywhere (the seedless algorithm). In stark contrast the NLO perturbative application of an algorithm with seeds has only the energetic partons themselves to act as seeds. Thus there is a dramatic change at NNLO where the extra parton can serve as a seed, as in Fig. 9, changing the found jet structure of the event even when the extra parton is quite low energy. This is the source of the perturbative Infrared sensitivity.

One of the main problems with the use of a seedless cone algorithm has been its slow speed with respect to the seeded cone algorithms. This has made its use in reconstruction of a large number of events difficult. Combined with the fact that, for inclusive distributions, the differences between the results from a seeded cone algorithm like midpoint (defined below) and a seedless algorithm tend to be on the order of a percent or less,[2] there was no strong motivation for its use. Recently, a new seedless algorithm (SISCone) [16] that has speeds comparable to the seeded cone algorithms has been developed, removing this difficulty. For this reason, the SISCone algorithm is being adopted by the experiments at both the Tevatron and LHC.[3] Note that the problems with dark towers and the smearing of stable solution points (discussed later in Section 3.4.1) still remain with a seedless algorithm.

To address the issue of seeds on the experimental side and the $R_{\rm sep}$ parameter on the phenomenological side, the Run II study [8] recommended using the midpoint cone algorithm, in which, having identified 2 nearby jets, one always checks for a stable cone with its center at the midpoint between the 2 found cones. Thus, in the imagery of Fig. 9, the central stable cone is now always looked for, whether there is an actual seed there or not. It was hoped that this would remove the sensitivity to the use of seeds and remove the need for the $R_{\rm sep}$ parameter. While this expectation is fully justified with the localized, short-distance configuration indicated in Fig. 9, more recent studies suggest that at least part of the difficulty with the missing stable cones at the midpoint position is due to the (real) smearing effects on the energy distribution in (y, ϕ) of showering and hadronization. Also it is important to note that, in principle, IR-safety issues due to seeds will reappear in perturbation theory at order NNNLO, where the midpoint is not

[1] The Run II recommendations [8] did include the suggestion of a seedless algorithm.

[2] In Ref. [16], a statement is made that the impact may be larger for some exclusive final-state observables.

[3] A *streamlined* (faster) version of the seedless algorithm was used during the early stages of CDF in Run II, but was dropped because of the near equivalence of the results obtained with the midpoint cone algorithm.

the only problem configuration (for example, a seed at the center of a triangular array of 3 hard and mergeable partons can lead to IR-sensitivity). Eliminating the use of seeds remains the most attractive option.

3.1.4. Merging

Before proceeding we must consider another important issue that arises when comparing the cone algorithm applied in perturbation theory with its application to data at the Tevatron. The practical definition of jets equaling stable cones does not eliminate the possibility that the stable cones can overlap, *i.e.*, share a subset (or even all) of their calorimeter towers. To address this ambiguity, experimental decisions had to be made as to when to completely merge the two cones (based on the level of overlap), or, if not merging, how to split the shared energy. Note that there is only a weak analogy to this overlap issue in the NLO calculation. As described in Fig. 6(a), there is no overlap in either the left-hand (1 cone $\rightarrow$ 1 jet) or right-hand (2 cones $\rightarrow$ 2 jets) regions, while in the middle (3 cones $\rightarrow$ 1 jet) region the overlap between the 3 cones is 100% and the cones are always merged. Arguably the phenomenological parameter $R_{\rm sep}$ also serves to approximately simulate not only seeds but also the role of noncomplete merging in the experimental analysis. In practice in Run I, CDF and DØ chose to use slightly different merging parameters. Thus, largely unknown to most of the theory community, the two experiments used somewhat different cone jet algorithms in Run I. The CDF collaboration cone algorithm, JETCLU [11], also employed another "feature" called ratcheting, that was likewise under-documented. Ratcheting ensured that any calorimeter tower in an initial seed was always retained in the corresponding final jet. Stability of the final cones was partially driven by this "no-tower-left-behind" feature. Presumably the two experiments reported compatible jet physics results in Run I due to the substantial ($\geq$10%) uncertainties. Note also that, after the splitting/merging step, the resulting cone jets will not always have a simple, symmetric shape in (y, ϕ), which complicates the task of correcting for the underlying event and leads to larger uncertainties. In any case the plan for Run II as outlined in the Run II studies [8], called for cone jet algorithms in the 2 collaborations as similar as possible. Unfortunately, during Run II the collaborations have evolved to employing somewhat different cone algorithms. On the merging question, CDF in Run II merges two overlapping cones when more than 75% of the smaller cone's energy overlaps with the larger jet. When the overlap is less, the overlap towers are assigned to the nearest jet. DØ, on the other hand, uses a criterion of a 50% overlap in order to merge. There is anecdotal evidence from several studies that a merging criterion of 75% may be "safer" in high-density final states [17], where there is a tendency for over-merging to occur. While it is not necessary that all analyses use the same jet algorithm, for purposes of comparison it would be very useful for the experiments to have one truly common algorithm. There is certainly a lesson to be learned here for the collaborations at the LHC.

3.1.5. Summary

In summary, the iterative cone algorithm benefits, in principle, from a simple geometric definition that allows simple corrections for the UE contributions. At the same time it suffers from the experimental use of seeds and the need to implement a split/merge stage. The inclusion of the $R_{\rm sep}$ parameter in the perturbative calculation, while allowing more detailed comparisons to data, has also served to confuse the situation. Before discussing the role of fragmentation and smearing in the cone algorithm, we turn now to the other primary jet algorithm, the k_T algorithm.

3.2. k_T algorithm

With the mixed history of success for the cone algorithm, the (as yet) less well-studied k_T algorithm [18,19] offers the possibility of more nearly identical analyses in both experiments and in perturbation theory. This algorithm, which was first used at electron–positron colliders, is based on a pair-wise recombination scheme intended, in some sense, to "undo" the splitting that occurs during the fragmentation stage. Two partons/particles/calorimeter towers are combined if their relative transverse momentum is less than a given measure. To illustrate the clustering process, consider a multi-parton final state. Initially each parton is considered as a proto-jet. The quantities

$$k_{T,i}^2 = p_{T,i}^2,$$
$$k_{T,(i,j)}^2 = \min(p_{T,i}^2, p_{T,j}^2)\frac{\Delta R_{i,j}^2}{D^2}$$

are computed for each proto-jet i and each pair of proto-jets ij, respectively. As earlier, $p_{T,i}$ is the transverse momentum of the ith proto-jet and $\Delta R_{i.j}$ is the distance (in y, ϕ space, $\Delta R_{i.j} = \sqrt{\left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2}$) between each pair of proto-jets. D is the parameter that controls the size of the jet (analogous to R_{cone}). If the smallest of the above quantities is a $k_{T,i}^2$, then that proto-jet becomes a jet and is removed from the proto-jet list. If the smallest quantity is a $k_{T,(i,j)}^2$, then the two proto-jets (i, j) are merged into a single proto-jet by summing their 4-vector components, and the two original entries in the proto-jet list are replaced by this single merged entry. This process is iterated with the corrected proto-jet list until all the proto-jets have become jets, *i.e.*, at the last step the $k_{T,(i,j)}^2$ for all pairs of proto-jets are larger than all $k_{T,i}^2$ for the proto-jets individually (*i.e.*, the remaining proto-jets are well-separated) and the latter all become jets.

Note that in the pQCD NLO inclusive k_T jet calculation, the parton pair with the smallest k_T^2 may or may not be combined into a single jet, depending on the $k_{T,i}^2$ of the individual partons. Thus the final state can consist of either 2 or 3 jets, as was also the case for the cone algorithm. In fact, the pQCD NLO result for the inclusive k_T jet cross-section [18] suggests near equality with the cone jet cross-section in the case that $D \simeq 1.35 R_{\text{cone}}$ (with no seeds, $R_{\text{sep}} = 2$). Thus the inclusive cone jet cross-section with $R_{\text{cone}} = 0.7$ ($R_{\text{sep}} = 2$) is comparable in magnitude to the inclusive k_T jet cross-section with $D = 0.9$, at least at NLO. In the NLO language illustrated in Fig. 6 the condition that the partons be merged in the k_T algorithm is that $z^2\left(d^2/D^2\right) < z^2$ or $d < D$. Thus, at NLO, the k_T algorithm corresponds to the cone algorithm with $R_{\text{cone}} = D$, $R_{\text{sep}} = 1$. The earlier result, $D \simeq 1.35 R_{\text{cone}}$ (with $R_{\text{sep}} = 2$), is just the NLO statement that the contribution of the rectangular region $0 \le d \le 1.35 R_{\text{cone}}$, $0 \le z \le 1$ is numerically approximately equal to the contribution of the rectangular region $0 \le d \le R_{\text{cone}}$, $0 \le z \le 1$ plus the (3 stable cone) triangular region $R_{\text{cone}} \le d \le (1 + z)\, R_{\text{cone}}$, $0 \le z \le 1$.

In contrast to the cone case, the k_T algorithm has no problems with overlapping jets and, less positively, every calorimeter tower is assigned to some jet. While this last result made some sense in the e^+e^- collider case, where every final-state particle arose from the short-distance process, it is less obviously desirable in the hadron collider case. While the k_T algorithm tends to automatically correct for the splash-out effect by re-merging the energy distribution smeared by showering and hadronization into a single jet, this same feature leads to a tendency to enhance the splash-in effect by "vacuuming up" the contributions from the underlying event and including them in the large $k_{T,i}^2$ jets. This issue is exacerbated when the luminosity reaches the point that

there is more than one collision per beam bunch crossing and pile-up is significant. This is now true at the Tevatron and will certainly be true eventually at the LHC. Thus while the (splash-out) fragmentation corrections for the k_T algorithm are expected to be smaller than for cone algorithms, the (splash-in) underlying event corrections will be larger. A test of our understanding of these corrections will be provided by the comparison of the D and $R_{\rm cone}$ parameter values that yield comparable experimental jet cross-sections. If we can reliably correct back to the fixed-order perturbative level for both the cone and k_T algorithms, we should see $D \simeq 1.35 R_{\rm cone}$. Note that this result assumes that the cone jet cross-section has been corrected to the value corresponding to $R_{\rm sep} = 2$. On the other hand, under-corrected splash-in contributions in the k_T algorithm will require $D < 1.35 R_{\rm cone}$ for comparable jet cross-section values (still assuming that $R_{\rm sep} = 2$ describes the cone results). If the cone algorithm jet cross-section has under-corrected splash-out effects ($R_{\rm sep} < 2$), we expect that an even smaller ratio of D to $R_{\rm cone}$ will be required to obtain comparable jet cross-sections (crudely we expect $D < (1 + 0.35(R_{\rm sep} - 1))R_{\rm cone}$ for $1 \leq R_{\rm sep} \leq 2$). As we will discuss below, the current studies at the Tevatron suggest that $D < R_{\rm cone}$ for comparable cross-sections implying that indeed the k_T algorithm is efficiently vacuuming up extra particles. However, once corrected back to the parton level, the k_T algorithm cross-section is smaller than the cone result for $D = R_{\rm cone}$ as expected.

Another concern with the k_T algorithm is the computer time needed to perform multiple evaluations of the list of pairs of proto-jets as 1 pair is merged with each pass, leading to a time that grows as N^3, where N is the number of initial proto-jets in the event. Recently [20] a faster version of the k_T algorithm, "Fastjet", has been defined that recalculates only an intelligently chosen sub-list with each pass and the time grows only as $N \ln N$, for large N. The software in [20] also provides alternative versions of the k_T algorithm. As defined above, the algorithm is the *inclusive* version, keeping all possible jets defined by the parameter D. There is also an *exclusive* version [19] where another parameter d_{cut} is introduced. Merging stops when all remaining $k^2_{T,(i,j)}$ and $k^2_{T,i}$ exceed d_{cut} and the remaining $k^2_{T,i}$ define the exclusively defined jets (*i.e.*, the previously removed $k^2_{T,i} < d_{cut}$ jets are discarded). The Fastjet code also includes the so-called Cambridge/Aachen k_T algorithm [21,22], the inclusive version of which is defined similarly to the algorithm above except that the prefactor $\min(p^2_{T,i}, p^2_{T,j})$ is absent from $k^2_{T,(i,j)}$ and $k^2_{T,i} = 1$. Finally the Fastjet code for the k_T algorithm also includes an innovative technique for defining the "area" of the jet and allowing a correcting for the UE contribution. Fake or "ghost" particles with exponentially tiny energies are added to each event on an essentially uniform grid in (y, ϕ) (*i.e.*, each ghost particle is representative of a fixed area in (y, ϕ)). The final jets found by the algorithm will then contain some of the ghost particles. While the kinematics of the jet is unchanged by the presence of the ghost particles (since they have such tiny energies), their number gives a measure of the area in (y, ϕ) of the jet.

It should also be noted that, although it would appear that the inclusive k_T algorithm is defined by a single parameter D, the suggested code for the k_T algorithm [23] includes several switches and parameters to fully define the specific implementation of the algorithm. Thus, as is the case for the cone algorithm, the k_T algorithm also exhibits opportunities for divergence between the implementation of the algorithm in the various experiments, and care should be taken to avoid this outcome.

3.3. Jet masses for jets at NLO

As has already been suggested, the invariant masses of jets are expected to play an increasingly important role at the LHC as a useful jet property. For example, jet masses can help to isolate

events where individual jets correspond to essentially all of the decay products of boosted heavy objects (say top quarks or W bosons). It is helpful to set the stage for these analyses by discussing first the expected magnitude of masses of jets arising from perturbative QCD interactions only. We will return to the question of jet masses in Section 6 where showering, hadronization and UE effects will be included. The NLO configurations with 2 partons in a jet, corresponding to $z > 0$ and $d > 0$, 1 cone $\to$1 jet and 3 cones $\to$ 1 jet in Fig. 6, will result in perturbative jets with nonzero masses. While the distribution of jet masses will be singular at the origin in perturbation theory, due to the soft and collinear singularities, the average jet mass (squared) is an infrared-safe quantity that can be evaluated order-by-order in perturbation theory. In particular, to find the average jet mass squared for a fixed jet p_J at NLO we simply evaluate the NLO inclusive jet cross-section at that fixed p_J-weighted by the corresponding jet mass squared, where the evaluation involves a sum over all 2-partons-in-a-jet configurations with the required p_J. We then divide this sum by the corresponding Born level inclusive jet cross-section. Note that the notation for the perturbative order used here corresponds to the order of the jet cross-section. Thus at LO jets correspond to single partons with vanishing jet mass. As noted above, NLO jets receive contributions from 2 parton configurations with nonzero jet masses, with the masses proportional to a single power of α_s. In detail this calculation is complicated due to the large number of perturbative processes that contribute and the fact that the available phase is restricted by the pdfs in a way that varies with $p_J/\sqrt{s}$. On the other hand, the expected general form of the perturbative jet mass is straightforward to motivate. By dimensional analysis the dominant contribution to the NLO jet mass squared will scale with p_J^2, will scale with the algorithm defined "size" of the jet, R^2 or D^2 and exhibit a factor of α_s. Choosing the factorization/renormalization scale in the running coupling to be $\mu = p_J/2$ we are led to expect for a cone algorithm

$$\left\langle M_J^2 \right\rangle_{\rm NLO} \simeq \overline{C} \left(\frac{p_J}{\sqrt{s}} \right) \alpha_s \left(\frac{p_J}{2} \right) p_J^2 R^2,$$

where the prefactor (prefunction) $\overline{C}$ has a magnitude of order unity and decreases slowly with increasing $p_J/\sqrt{s}$. Using the CTEQ 6.2 pdfs, averaging over jet rapidities in the range $|y_J| \leq 2.5$ and employing the EKS NLO inclusive jet code [10] for the cone jet algorithm, we find the average NLO jet masses illustrated in the following figures. Due to the expected simple linear momentum dependence we will focus on the linear jet mass, $\sqrt{\langle M_J^2 \rangle_{\rm NLO}} \simeq \sqrt{\overline{C}\alpha_s}\, p_J R$. The simplest dimensionful NLO predictions for the jet mass at both Tevatron and LHC energies are illustrated in Fig. 10.

We see that for low momentum jets, $p_J \sim 100$ GeV, the jet mass is relatively independent of the overall energy s. On the other hand, for substantially larger momenta the reduced phase space at the Tevatron leads to smaller predicted jet masses at the same momentum. We can simplify the discussion by first scaling out the (necessary dimensionful) linear dependence on p_J as displayed in Fig. 11. As suggested earlier we can largely correct for the overall energy difference by plotting instead versus the momentum *fraction*, $x_T = 2p_J/\sqrt{s}$, as illustrated in Fig. 12. Now the impact of the differing beam energies (and beam flavor) is much reduced with scaled jet mass distributions of very similar shape, and the scaled LHC jet mass smaller than the scaled Tevatron jet mass at the same x_T by approximately 10%–15%. The common falling shape of the distributions as functions of x_T can be understood as arising from the falling coupling, the falling pdfs and the transition from dominantly gluons at small x_T with $\overline{C} \propto C_A = 3$ to dominantly quarks (and anti-quarks) at large x_T with $\overline{C} \propto C_F = 4/3$. The difference in the magnitude of the two distributions arises from the scale breaking in the theory (p_J is larger at

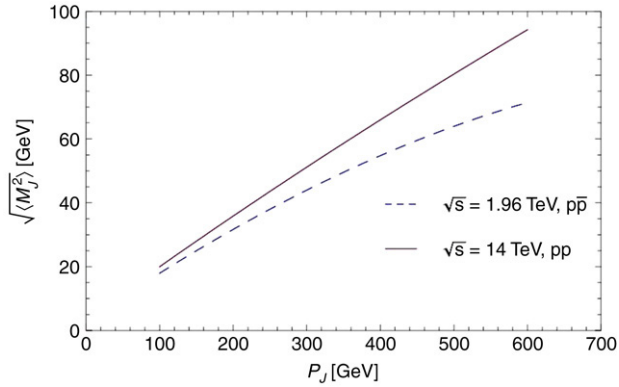


Fig. 10. NLO jet masses for a cone jet with $R = 0.7$ and $R_{\text{sep}} = 1.3$ with Tevatron and LHC energies and beams.

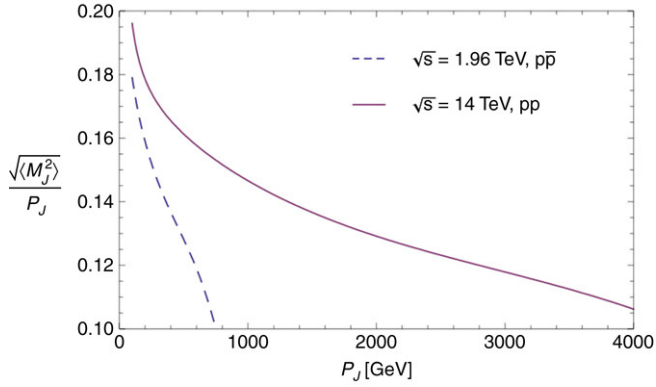


Fig. 11. NLO jet mass for a cone jet with $R = 0.7$ and $R_{\text{sep}} = 1.3$ scaled by the jet momentum p_J with Tevatron and LHC energies and beams.

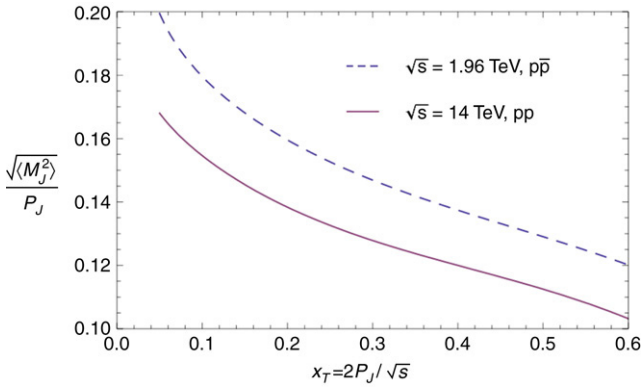


Fig. 12. NLO jet mass for a cone jet with $R = 0.7$ and $R_{\text{sep}} = 1.3$ scaled to the jet momentum p_J plotted versus $x_T = 2p_J/\sqrt{s}$ with Tevatron and LHC energies and beams.

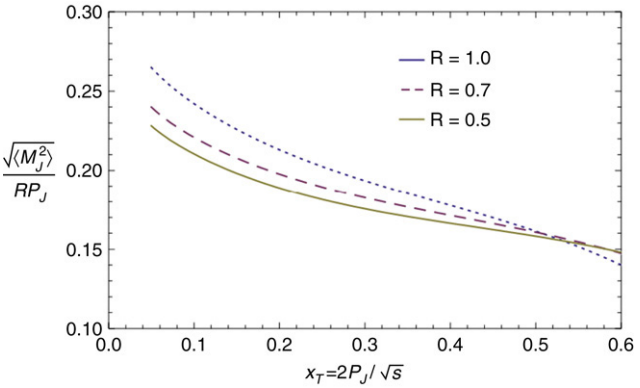


Fig. 13. NLO jet mass for a cone jet with $R_{\text{sep}} = 1.3$, $\sqrt{s} = 14$ TeV scaled by the jet momentum p_J and the cone radius R plotted versus $x_T = 2p_J/\sqrt{s}$ for various values of the cone radius R.

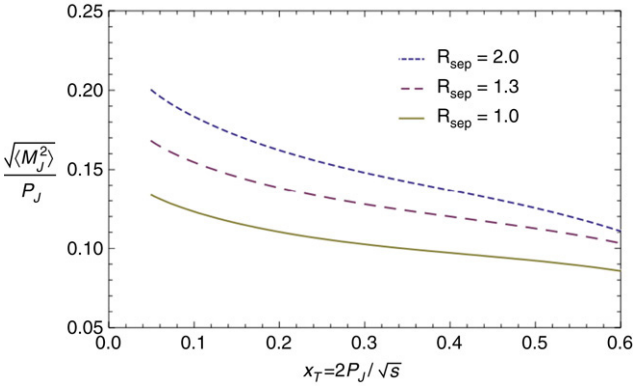


Fig. 14. NLO jet mass for a cone jet with $R = 0.7$, $\sqrt{s} = 14$ TeV scaled by the jet momentum p_J plotted versus $x_T = 2p_J/\sqrt{s}$ for various values of the cone radius R_{sep}.

the LHC at the same x_T) with both the running coupling and the running pdfs being smaller at the LHC. We can verify the approximately linear dependence on the cone radius R by both varying the radius and scaling it out as in Fig. 13 for LHC beam energies.

While the dependencies on p_J and R are not exactly linear, the NLO jet mass is remarkably well-described by the simple rule-of-thumb

$$\sqrt{\langle M_J^2 \rangle_{\text{NLO}}} \approx 0.2 p_J R,$$

where the numerical prefactor of 0.2 (approximately) includes the dependence on $\sqrt{\alpha_s}$, the color charges (a mix of C_F and C_A) and the pdfs. For better than 25% accuracy the still simple prefactor $(0.2 + (0.3 - x_T)/6)$ suffices.

Finally consider the dependence on the *ad hoc* parameter R_{sep}. This is illustrated in Fig. 14. We see the expected monotonic dependence on R_{sep}, where the variation is somewhat less rapid than linear. In the language introduced earlier we see that the mass for a naive Snowmass cone jet ($R_{\text{sep}} = 2.0$) is expected (in NLO perturbation theory) to be approximately 30%–50% larger than the mass of a k_T jet ($R_{\text{sep}} = 1.0$) with $D = R$. Qualitatively we anticipate that, compared to

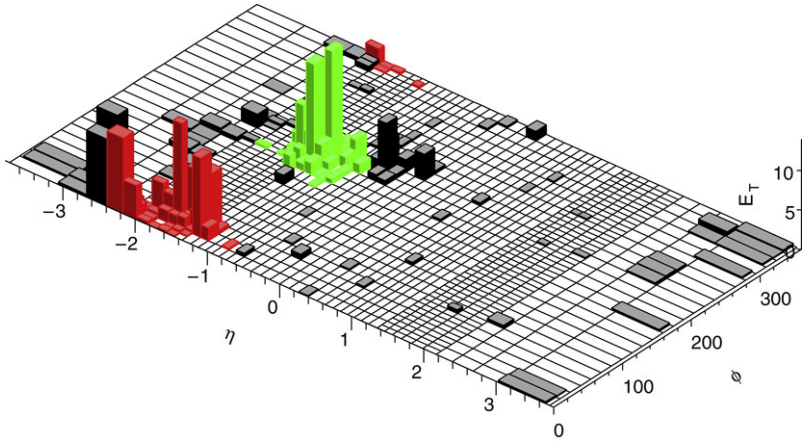


Fig. 15. An example of a Monte Carlo inclusive jet event where the midpoint algorithm has left substantial energy unclustered.

NLO perturbation theory, the inclusion of showering, hadronization and the underlying event will lead to "splash-out" effects and smaller jet masses for the cone algorithm, but "splash-in" and larger jet masses for the k_T algorithm. Hence the masses of jets found in realistic environments by different algorithms, as discussed in Section 6, are expected to be more similar than suggested by NLO perturbation theory and Fig. 14.

In summary, perturbative QCD (alone) leads us to expect jet masses at the LHC that grow slightly more slowly than linearly with the jet momentum, scale linearly with the jet algorithm defined jet size, R or D, and have a magnitude approximately ($\pm 25\%$) described by $\sqrt{\langle M_J^2 \rangle_{\rm NLO}} \approx 0.2 p_J R$. The explicit prefactor (0.2) corresponds to $R_{\rm sep} = 1.3$ cone jets with Snowmass jets ($R_{\rm sep} = 2.0$) having masses larger by 10% and NLO k_T jets ($R_{\rm sep} = 1.0$) having masses smaller by 20%.

3.4. Recent cone algorithm issues

3.4.1. Jets at the "smeared" parton level and dark towers

In studies of the Run II iterative cone algorithms a previously unnoticed problem has been identified [24] at the particle and calorimeter levels, which is explicitly not present at the NLO parton level. It is observed in a (relatively small) fraction of the events that some energetic particles/calorimeter towers remain unclustered in any jet. This effect is understood to arise in configurations of two nearby (*i.e.*, nearby on the scale of the cone size) showers, where one shower is of substantially larger energy. Any trial cone at the location of the lower energy shower will include contributions from the larger energy shower, and the resulting centroid will migrate towards the larger energy peak. This feature is labeled "dark towers" in Ref. [24], *i.e.,* clusters that have a transverse momentum large enough to be designated either a separate jet or to be included in an existing nearby jet, but which are not clustered into either. A Monte Carlo event with this structure is shown in Fig. 15, where the towers unclustered into any jet are shaded black.

A simple way of understanding the dark towers can be motivated by returning to Fig. 7, where the only smearing in (y, ϕ) between the localized energy distribution of panel (a) (the individual partons) and the "potential" of panel (b) arises from the size of the cone itself. On the other

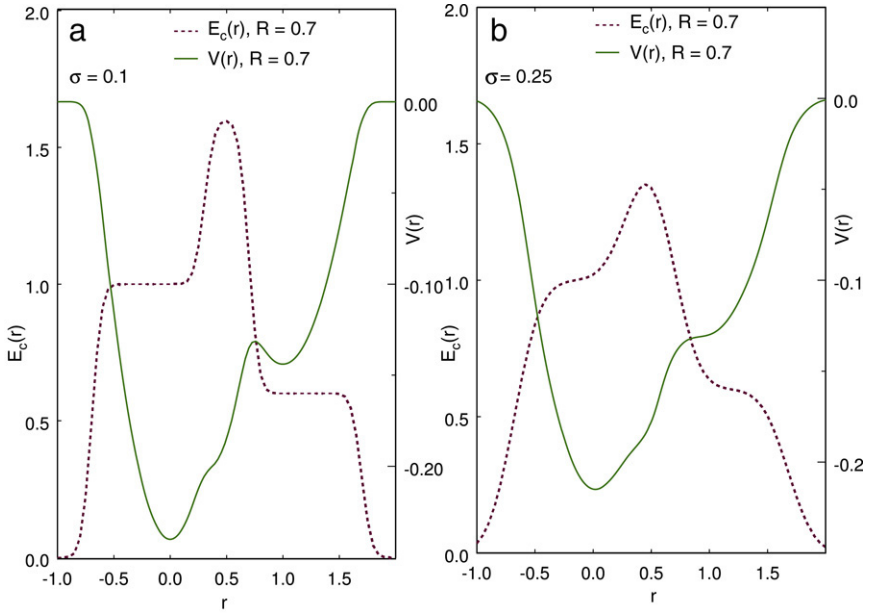


Fig. 16. Energy-in-cone and potential distributions corresponding to Gaussian smearing with (a) $\sigma = 0.1$ and (b) $\sigma = 0.25$ for $d = 1.0$ and $z = 0.6$.

hand, we know that showering and hadronization will lead to further smearing of the energy distribution and thus of the potential. Sketched in Fig. 16 is the potential (and the energy-in-cone) distributions that results from Gaussian smearing with a width of (a) $\sigma = 0.1$ and (b) $\sigma = 0.25$ (in the same angular units as $R = 0.7$). Note that the smearing of the potential occurs in 2 dimensions and that here we are considering only a 1-dimensional slice. In both panels, as in Fig. 7, the partons have p_T ratio $z = 0.6$ and angular separation $d = 1.0$. Note that as the smearing increases from zero as in panel (b) of Fig. 7 to the smeared results in Fig. 16, we first lose the (not so deep) minimum corresponding to the midpoint stable cone (and jet), providing another piece of the explanation for why showers more than $1.3 \cdot R_{\text{cone}}$ apart are not observed to merge by the experiments. In panel (b), with even more smearing, the minimum in the potential near the shower from the lower energy parton also vanishes, meaning this (lower energy) shower is part of no stable cone or jet, *i.e.*, leading to dark towers. Any attempt to place the center of a trial cone at the position of the right parton will result in the centroid "flowing" to the position of the left parton and the energy corresponding to the right parton remaining unclustered in any jet. Note that the Run I CDF algorithm, JETCLU with ratcheting, limited the role of dark towers by never allowing a trial cone to leave the seed towers, the potential dark towers, behind. The effective smearing in the data is expected to lie between σ values of 0.1 and 0.25 (with shower-to-shower fluctuations and some energy dependence, being larger for smaller p_T jets) making this discussion relevant, but this question awaits further studies. Note that Fig. 16 also suggests that the midpoint algorithm will not entirely fix the original issue of missing merged jets. Due to the presence of (real) smearing this middle cone is often unstable and the merged jet configuration will not be found even though we explicitly look for it with the midpoint cone. Thus, even using the recommended midpoint algorithm (with seeds), there may remain a phenomenological need

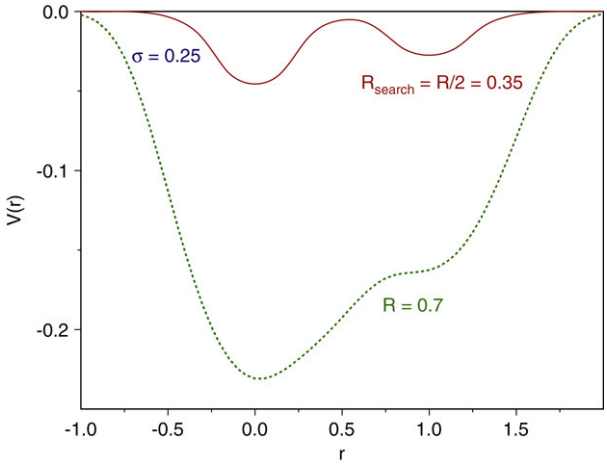


Fig. 17. The stable cone finding potential with the reduced search cone, $R_{\text{search}} = R_{\text{cone}}/2$. The original potential from Fig. 16, panel (b) with $R_{\text{search}} = R_{\text{cone}}$ is indicated as the dashed curve.

for the parameter value $R_{\text{sep}} < 2$, *i.e.*, a continuing mismatch between data and pQCD that requires correction.

3.4.2. The search cone algorithm

A potential solution for the dark towers problem is described in Ref. [24]. The idea is to decouple the jet finding step from the jet construction step. In particular, the stable cone finding procedure is performed with a cone of radius half that of the final jet radius, *i.e.*, the radius of the search cone, $R_{\text{search}} = R_{\text{cone}}/2$. This procedure reduces the smearing in Figs. 7 and 16, and reduces the phase space for configurations that lead to dark towers (and missing merged jets). This point is illustrated in Fig. 17, which shows the potential of Fig. 16, panel (b) corresponding to the reduced radius search cone. Note, in particular, that there is again a minimum at the location of the second parton. Seeds placed at each parton will yield a stable cone at each location even after the smearing. Using the smaller search cone size means there is less influence from the (smeared) energy of nearby partons. After identifying the locations of stable cones, the larger cone size, *e.g.*, $R_{\text{jet}} = R_{\text{cone}} = 0.7$, is used to sum all objects inside and construct the energy and momentum of the jet (with no iteration). All pairs of stable cones separated by less than $2R_{\text{cone}}$ are then used to define midpoint seeds as in the usual midpoint cone algorithm. A trial cone of size R_{cone} is placed at each such seed and iterated to test for stability. (Note that this midpoint cone is iterated with cone size R_{cone}, not the smaller R_{search}, contrary to what is described in the literature [25].) Thus, just as in the midpoint cone algorithm, stable midpoint cones will be found by the CDF search cone algorithm. However, as already discussed, we expect that there will be no stable midpoint cone due to the smearing. Note that, even with the reduced smearing when using the smaller search cone radius, there is still no central stable cone in the potential of Fig. 17. On the other hand, as applied to NLO perturbation theory without smearing, the search cone algorithm should act like the usual midpoint cone algorithm and yield the naïve result of Fig. 6(a). The net impact of adding the step with the smaller initial search cone as applied to data is an approximately 5% increase in the inclusive jet cross-section. In fact, as applied to data the search cone algorithm identifies so many more stable cones, that the CDF collaboration has

decided to use the search cone algorithm with the merging parameter $f_{\text{merge}} = 0.75$ (instead of 0.5) to limit the level of merging.

Unfortunately, a disturbing feature of the search cone algorithm arises when it is applied at higher orders in perturbation theory as was pointed out during the TeV4LHC Workshop. At NNLO in perturbation theory, the search cone algorithm can act much like the seeds discussed earlier. In particular, the search cone algorithm can identify a (small radius R_{search}) stable (soft) cone between two energetic cones arising from the soft gluon between 2 energetic partons configuration discussed earlier. The soft search cone is stable exactly because it "fits" between the two energetic partons without including either; the spacing between the two energetic partons can be in the range $2R_{\text{search}} = R_{\text{cone}} < \Delta R < 2R_{\text{cone}}$. Then, when the radius of the (stable, soft) search cone is increased to R_{cone}, the resulting full size cone will envelop, and serve to merge, the two energetic partons. This can occur even when the two energetic partons do not constitute a stable combined cone in the standard cone algorithm. Thus at NNLO, the search cone algorithm can exhibit an IR-sensitivity very similar to, and just as undesirable as, the seed-induced problem discussed earlier. The conclusion is that the search cone algorithm, while it does address the dark tower issue, creates its own set of issues and is not considered to be a real solution of the dark tower problem.

3.4.3. The midpoint cone algorithm with a "second pass"

The recommendation adopted in the TeV4LHC workshop, and which we also endorse, is to remove the search cone from the midpoint cone algorithm and to substitute a "second pass" step. After all stable cones are found in an event, the towers corresponding to those jets are removed from the list, and the midpoint cone algorithm is run again. The "dark towers" are then reconstructed as jets in their own right, as the attractive influence of nearby larger jets is no longer present. The question remains as to what to do with the second pass jets. According to the criteria discussed above, some of them should be merged with the larger nearby jet, while others should remain as separate jets. As separate jets, the second pass jets have an insignificant impact on inclusive jet cross-sections. If added to the nearby larger jets, they can have an effect of the order of up to 5%. The correct treatment is still under investigation. Note that the second pass jets may also be important in the accurate reconstruction of a complicated multi-jet final state, such as in $t\bar{t}$ events.

It is also important to keep in mind that the experimental reality in which jet algorithms are applied is more complex than the 1-dimensional slice shown in Fig. 16. The final-state hadrons are distributed in two dimensions (y and ϕ) and the resultant Snowmass potential has fluctuations in these two dimensions. This has implications regarding the presence or not of dark towers and of the effect of the distribution of the jet energies on the stability of cone centroids.

3.4.4. Summary

In summary, to compare fixed-order perturbation theory with data there must be corrections for detector effects, for the splash-in contributions of the underlying event (and pile-up) and for the splash-out effects of hadronization (and for showering when comparing to leading-order calculations). It is the response to both the splash-in and splash-out effects that distinguishes the various cone algorithms and drives the issues we have just been discussing. It is also important to recognize that the splash-in and splash-out effects (and corrections) come with opposite signs and there can be substantial cancellation. We will return to the question of Run II corrections below. The conclusion from the previous discussion is that it would be very helpful to include also a correction in the experimental analysis that accounts for the use of seeds, or to use a seedless

cone algorithm such as SISCone [16]. Then these experimental results could be compared to perturbative results without seeds, avoiding the inherent infrared problems caused by seeds in perturbative analyses. At the same time, the analysis described above suggests that using the midpoint cone algorithm, to remove the impact of seeds at NLO, does not fully eliminate the impact of the smearing due to showering and hadronization, which serves to render the midpoint cone of fixed-order perturbation theory unstable. The same may be true, but possibly to a lesser extent, with the use of a seedless algorithm such as SISCone, which increases the number of identified stable cones. Thus we should still not expect to be able to compare data to NLO theory with $R_{\rm sep} = 2$, although the impact may be somewhat reduced. The possible downside of finding more stable cones is that the split/merge step will play a larger role, with no analogue effect in the perturbative analysis. Splitting the calorimeter towers with a host of low E_T stable cones can serve to lower the E_T of a leading jet and introduce an undesirable dependence on the minimum p_T cut defining which stable cones are included in the split/merge process. This is surely only a percent level correction, but control at that level is the ultimate goal.

3.5. Jets at the hadron level

The last step before the detector in our factorized, approximate picture of a hard scattering event is the hadronization step. Here the colored partons arising from the hard scattering itself, from ISR and FSR and from the UE are organized into color-singlet hadrons. Thus this step requires information about the color flow within the event as the event evolves from short to long distance, and that information must be maintained during the showering process. The initial color single hadrons will include resonances, which are then decayed into the long lifetime, long-distance final hadrons (primarily pions and kaons) according to the known properties of the resonances. Essentially by definition, this step represents nonperturbative physics, is model dependent and therefore not well-understood. On the other hand, by construction, it looks for minimal mass color-singlet final states while conserving energy and momentum and so results in only a minimal kinematic rearrangement. Thus, although there is considerable uncertainty about the details of hadronization, the impact on jet reconstruction is small and is fairly easily corrected for in both types of algorithms except perhaps at the lowest jet energies. In experimental analyses, these corrections are estimated using leading-order parton-shower Monte Carlo generators, by observing the variation of the predicted jet cross-sections after turning off the interaction between beam remnants and the hadronization. This procedure relies on the Monte Carlo providing a good description of those observables in the data that are most sensitive to nonperturbative contributions such as, for example, the underlying event energy away from jets and the internal structure of the jets. Recent precise measurements on jet shapes [25], as indicated in Fig. 18, have allowed the detailed study of the models employed to describe the underlying event in inclusive jet production at the Tevatron (see also Ref. [26]). Besides the contribution from the UE the jet shape is also sensitive to the changing character of the scattered parton, *i.e.*, gluons at low p_T and quarks at high p_T, and to the perturbative scaling of the jet size with $\alpha_s\,(p_T)$. The Tevatron studies, which will be discussed in more detail below, suggest that these last effects are reasonably well-described both by the showers in Monte Carlo simulations and by pQCD analyses. Future measurements of the underlying event in Run II, for different hadronic final states, promise to play a major role in the understanding of the measured jet cross-sections at the LHC.

Overall the challenge for the future, as noted above, is to continue to reduce the systematic uncertainties below the current 10%–50% level, which effectively guarantees agreement with

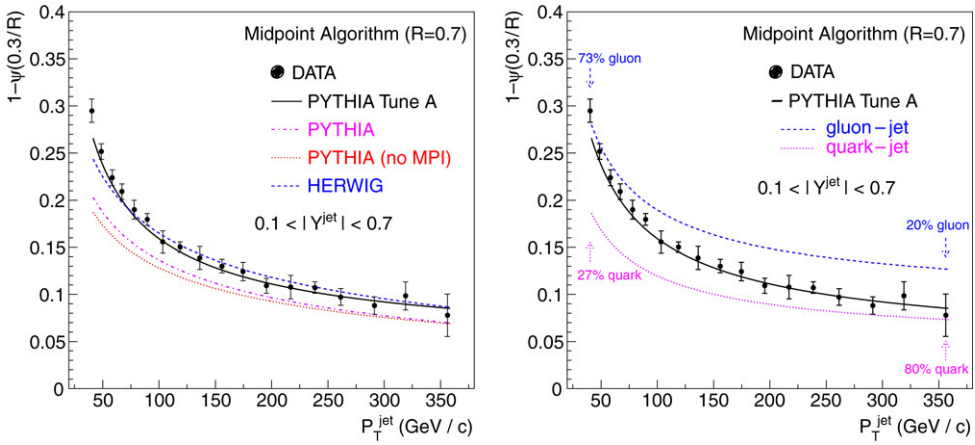


Fig. 18. The fraction of p_T in a cone jet of radius 0.7 that lies in the annulus from 0.3 to 0.7 as a function of jet p_T. The CDF Run II measurement is compared to predictions from Herwig and Pythia with different set of parameters (left), and to the separate predictions for quark and gluon jets (right). Figures from Ref. [25].

theory. Indeed, Run II studies of the corrections due to splash-in, *i.e.*, the underlying event (and pile-up), and the splash-out corrections due to hadronization are much more sophisticated than in Run I and presented in such a way that they can be applied either to the data (corrected for the detector) or to a theoretical (perturbative) calculation. The evaluation of these corrections is based on data and the standard Monte Carlos, Pythia and Herwig, especially Tune A of Pythia, which seems to accurately simulate the underlying event in terms of multiple parton interactions, including the observed correlations with the hard scattering process [26]. We turn now to a detailed discussion of the experiences and lessons learned at the Tevatron.

4. Jets at the Tevatron

Most of the interesting physics processes in $p\bar{p}$ collisions at the Tevatron include jets in the final state. Therefore, the reconstruction and precise measurement of jets serve as key ingredients for the better understanding of hard scattering processes and for new physics searches at the Tevatron. As an example, consider top quark production. The top quark was discovered by the CDF and DØ experiments in 1995 [27–29], and since then studies of the top quark production and properties have been considered to be one of the primary physics goals for the Tevatron experiments. Top quark decays always have at least one jet in the final state and the systematic uncertainty of the top quark mass measurement is dominated by the jet energy scale uncertainty. The search for the Standard Model Higgs boson is also an important aspect of the Tevatron physics program. The Standard Model Higgs boson dominantly decays to $b\bar{b}$ for masses $m_H \lesssim 135$ GeV/c. The $b\bar{b}$ di-jet mass resolution has been stressed as one of the critical factors in the search for a light Higgs boson decay at the Tevatron [30,31]; a 20% decrease of $\sigma_m/m_{b\bar{b}}$ is expected to be have a similar effect as a 20% increase in the accumulated luminosity. Inclusive jet cross-section measurements have also been undertaken at the Tevatron. Such measurements are sensitive to new physics such as quark compositeness and the presence of new heavy particles from Beyond-the-Standard-Model scenarios, and also they provide crucial information on parton distribution functions for the proton.

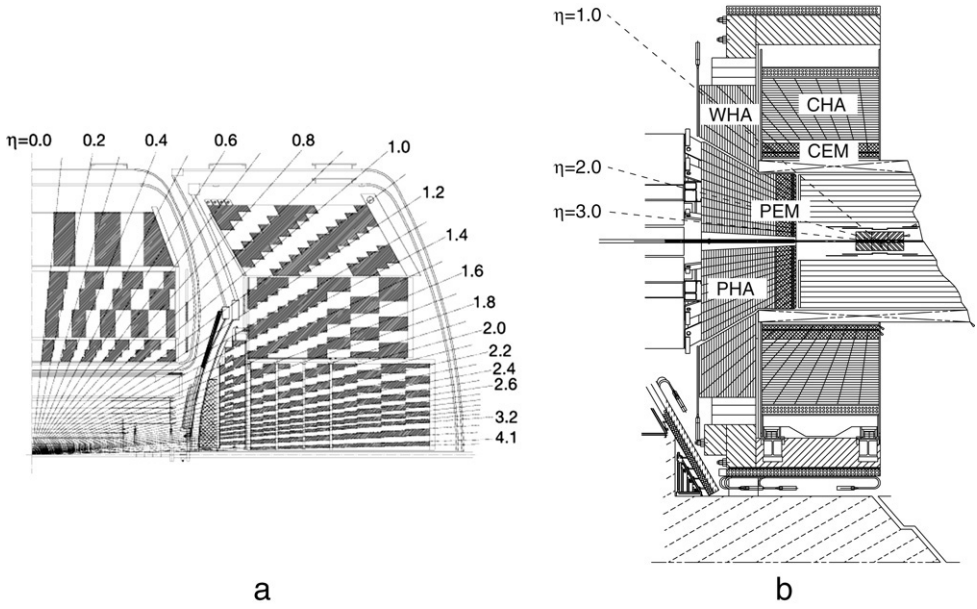


Fig. 19. Schematic views of a portion of the DØ (a) and CDF (b) calorimeters. Figures from Refs. [32,33], respectively.

4.1. Detectors

CDF and DØ are the two general-purpose detectors designed for measurement of $p\bar{p}$ collisions at the Fermilab Tevatron collider. Both detectors are composed of a solenoidal-magnet charged particle spectrometer surrounded by (sampling) calorimetry and a system of muon chambers. The components most relevant for jets are the calorimeters, which are used to measure the energy and angle of particles produced in $p\bar{p}$ collisions. The CDF and DØ calorimeters both have a projective tower geometry as shown in Fig. 19.

The DØ calorimeters are uranium and liquid argon sampling calorimeters. The DØ calorimetry consists of a central section covering $|\eta| \lesssim 1$ and an endcap section extending the coverage to $|\eta| \sim 4$. The calorimetry has three sections: EM, fine hadronic (FH) and coarse hadronic (CH). In the CH section, copper or steel is used instead of uranium as an absorber. The DØ calorimeters are nearly compensating, with an e/π ratio less than 1.05 above 30 GeV. The tower segmentation in η–ϕ space is $0.1 \times 5.625°$. The average noise per channel is on the order of 50 MeV. The relatively long charge collection time for the liquid argon signal and the resultant electronic shaping of the signal results in the possibility of towers with (modest) negative energy.

The CDF electromagnetic (EM) section consists of alternating layers of lead and scintillator, while the hadronic (HAD) section consists of alternating layers of steel and scintillator. The CDF calorimetry is divided into two main pseudorapidity ($|\eta|$) regions; the central calorimeter covers $|\eta| < 1.1$ and the plug region covers $1.1 < |\eta| < 3.6$. The region between the central and plug regions is covered by the end-wall hadron calorimeter which has a similar construction as the central hadron calorimeter. The tower size in the central region is $\Delta\eta \times \Delta\phi \approx 0.1 \times 15°$, and the segmentation in the plug region varies as a function of η ($\Delta\eta \times \Delta\phi \approx 0.1 \times 7.5°$ for $|\eta| < 1.8$ and with $\Delta\eta$ increasing for larger $|\eta|$ values). The noise level is very low, with only ∼1 noise

tower with $E_T > 50$ MeV being present per event. The E_T threshold on each calorimeter tower used for jet reconstruction is 100 MeV, and no special care is necessary for noise suppression.

4.2. Jet reconstruction and energy measurement

In contrast to leptons or photons, jets are clusters of "objects" that have to be defined by a clustering algorithm, as discussed previously in this paper. In parton-shower Monte Carlo events, a jet clustering algorithm can be applied to particles produced after hadronization of the partons, producing hadron-level jets; for pQCD parton-level predictions, jets are formed by running the algorithm on partons from the fixed-order (typically next-to-leading-order) pQCD event generator. Experimentally, it is extremely challenging to both identify and measure the kinematic properties of the large variety of individual particles produced in high-energy hadronic collisions, especially with the relatively coarse calorimeter segmentation present in the Tevatron experiments. As a result jets are reconstructed by running the jet algorithm on the "raw" energy depositions in the calorimeter towers, and not on the individual particles.

The reconstruction and energy measurement of these "calorimeter" jets are affected by a variety of instrumental and physics effects. The instrumental effects include (1) calorimeter nonuniformity, (2) resolution effects due to large fluctuations in particle showering in the calorimeter, (3) nonlinear response of the calorimeter, especially to hadrons, and (4) low momentum particles not reaching the calorimeter due to materials in front of the calorimeter and the solenoidal-magnetic field. Energy from additional $p\bar{p}$ interactions occurring in the same bunch crossing also affects the jet reconstruction and energy measurement. This has a nonnegligible effect especially for low transverse momentum jets and high instantaneous luminosities. The situation is further complicated due to the fact that (a) the underlying event contributes energy to the jet cone, and that (b) the jet cone does not contain all the energy of the parent parton because of the effects of parton showering and hadronization. All of these effects have to be taken into account when comparing experimental data with theoretical predictions or extracting physics quantities of interest from experimental data.

Analyses take these physics effects into account in different manners. For example, in new physics searches or in measurements of top quark properties, attempts are often made to correct the jets measured in the calorimeters back to the parent parton from which the jet originated. Extraction of the physics quantities of interest is done by comparing the data, after background subtraction, with signal events implemented in the parton-shower Monte Carlo.

Typically in jet cross-section measurements, the data are corrected to the hadron-level (*i.e.*, corrected only for instrumental effects); at this level, the results can be compared with hadron-level theoretical predictions without any knowledge of the detectors. We strongly encourage that, where possible, measurements at the Tevatron or LHC produce results at this level; often, instead the predictions from Monte Carlos are themselves passed through a detector simulation program and then directly compared to the "raw" data distributions.

Sometimes, the theoretical predictions from perturbative QCD calculations are available only at the parton level and then it is necessary to either correct the theory to the hadron level or the data to the parton level. For inclusive jet cross-sections, the best theoretical predictions available, as of now, are from next-to-leading-order perturbative QCD calculations [10,34,35]. When making comparisons between experimental measurements and next-to-leading (NLO) - order predictions, nonperturbative QCD effects from the underlying event and hadronization (see Section 4.3 for details) are evaluated based on tuned parton-shower Monte Carlos. Note that a correction is not made for "out-of-cone" energy from hard gluon emission, as such effects are

already (at least partially) taken into account in the NLO calculations. In these comparisons, next-to-leading-order hard gluon emission is assumed to model the whole parton-shower perturbative QCD radiation process. NLO pQCD calculations have been shown to provide a reasonable description of energy flow inside jets [36], so this is considered to be a reasonable assumption[4]; however, data–theory comparisons will benefit a great deal from higher-order QCD calculations or next-to-leading matrix element calculations interfaced with parton-shower models, as in *e.g.*, MC@NLO [37].

The jet algorithms and procedures for jet energy scale correction to the hadron-level (*i.e.*, correction for instrumental effects) employed by CDF and DØ are discussed below.

4.2.1. Jet algorithms

In Run II, DØ uses the midpoint iterative cone algorithm with the cone radius $R_{\text{cone}} = 0.5$ and 0.7 and merging criterion (see Section 3.1.4) $f_{\text{merge}} = 0.50$. At CDF, several jet algorithms are in use. In a large fraction of analyses, the JETCLU cone clustering has been used, which was used also in Run I at CDF, with the cone radius $R_{\text{cone}} = 0.4$ and 0.7 and the merging fraction $f_{\text{merge}} = 0.75$. The midpoint algorithm with the search cone step, *i.e.*, the search cone algorithm (see Section 3.4.2), with the cone radius of $R_{\text{cone}} = 0.7$ and the merging fraction of $f_{\text{merge}} = 0.75$ was also used in several jet cross-section measurements; however, because of the IR-sensitivity introduced by the search cone step, the search cone step is being removed from the clustering. The k_T algorithm was also used in the inclusive jet cross-section measurements with $D = 0.5$, 0.7 and 1.0.

4.2.2. Jet energy scale at DØ

The jet energy calibration procedure employed by DØ is based primarily on data, exploiting the conservation of transverse momentum [38]. The measured jet energy is corrected back to the true hadron-level jet energy by:

$$E_{\text{jet}}^{\text{hadron}} = \frac{E_{\text{jet}}^{\text{measured}} - E_0}{R_{\text{jet}} \cdot S}, \tag{1}$$

where E_0 is an offset energy which includes the uranium noise, energy from the previous bunch crossing, and additional $p\bar{p}$ interactions. The offset E_0 is determined by measuring the transverse energy density in the minimum-bias and zero-bias data. R_{jet} represents the calorimeter response to jets. This corrects both for the calorimeter nonuniformity in η and for the absolute energy scale. S is the showering correction, which corrects for the energy emitted outside the jet cone due to detector effects.

DØ uses the missing E_T projection fraction (MPF) method [38], which exploits the transverse momentum conservation in an event, to measure the calorimeter response to jets (R_{jet}). In photon + jet events, for example, the transverse energies of the real photon and the other recoil particles at the hadron level should satisfy:

$$\vec{E}_T^{\,\gamma} + \vec{E}_T^{\,\text{recoil}} = 0. \tag{2}$$

In general, the calorimeter response to both photons and recoils is less than unity and the energy mis-measurement results in missing E_T ($\not\!\!E_T$) in events:

[4] As $r \to 0$ inside a jet, however, the energy profile function develops large logarithms of the form $\alpha_s^n \log^{2n-1} r$; for an accurate description of this region, these terms would have to be re-summed. The pure NLO description should be accurate for r values greater than 0.3.

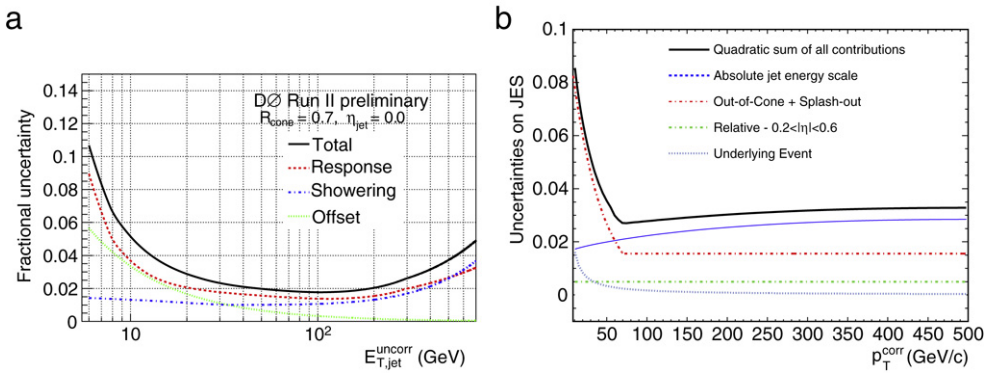


Fig. 20. The jet energy scale systematic uncertainties for (a) central jets at $\eta = 0$ as a function of uncorrected jet energy at DØ and (b) central jets in $0.2 < |\eta| < 0.6$ as a function of corrected jet p_T at CDF. Figures from Refs. [32,33].

$$R_\gamma \vec{E}_T^{\,\gamma} + R_{\text{recoil}} \vec{E}_T^{\,\text{recoil}} = -\not\vec{E}_T. \tag{3}$$

After the EM energy calibration, $R_\gamma = 1$, and Eqs. (2) and (3) transform to:

$$R_{\text{recoil}} = 1 + \frac{\not\vec{E}_T \cdot \vec{n}_T^{\,\gamma}}{E_T^\gamma}, \tag{4}$$

where $\vec{n}_T^{\,\gamma} = \vec{E}_T^{\,\gamma}/|\vec{E}_T^{\,\gamma}|$. In back-to-back photon + jet events, R_{recoil} can be considered as the response of a jet (R_{jet}). The absolute jet response correction is determined and applied after the response is equalized as a function of η. Fig. 20(a) shows the jet energy scale systematic uncertainty achieved for central jets as a function of jet energy at DØ.

4.2.3. Jet energy scale at CDF

At CDF, the determination of the jet energy absolute (response) correction at the hadron-level relies primarily on a detector simulation and jet fragmentation model [33]. Therefore, the accuracy of the calorimeter simulation is crucial for the precise jet energy scale determination. The CDF calorimeter simulation response to single particles is tuned to reproduce the response measured both in the test beam and in the collision data taken with the real detector. The CDF detector simulation uses a Geant-based detector simulation in which a parametrized shower simulation (Gflash [39]) is used for the calorimeter response. The use of Gflash was primarily motivated by its excellent performance in terms of speed, but also the relative ease of tuning. After tuning, the absolute correction for calorimeter-level jets to the corresponding hadron-level jets is obtained on average from di-jet Monte Carlo events by matching the two leading hadron-level jets to the corresponding calorimeter jets, and then taking the most probable value of calorimeter-jet p_T for hadron-level jets with a given $p_{T,\text{jet}}^{\text{hadron}}$.

This method can be used in the entire calorimeter coverage region; however, in practice the tuning is limited in precision outside the central region ($0.2 < |\eta| < 0.6$) because of the limited tracking coverage. Thus, the jet energy scale outside the central region is rescaled to that of the central region based on the p_T balance of the leading two jets in exclusive (*i.e.*, stringent cuts are placed on the presence of any additional jets in the event) di-jet events; in such di-jet events, two jets should have the same p_T to leading order and any imbalance is due to calorimeter nonuniformity. This correction is called the relative correction.

Additional $p\bar{p}$ interactions, in the same bunch crossing as the interaction which produced the jets, also contribute energy in the jet cone. The number of reconstructed primary vertices ($N_{\rm vtx}$) is a good estimator of the number of interactions in the same bunch crossing. For jets reconstructed with the cone algorithm, this multiple interaction correction, p_T^{MI}, is derived by measuring the transverse momentum in a cone of the same size as the jet cone in the central region as a function of $N_{\rm vtx}$ in minimum-bias events. For a cone size $R_{\rm cone} = 0.7$, the correction subtracts $p_T^{MI} \sim 1$ GeV/c for each additional primary vertex.

The CDF jet energy corrections described above can be summarized as,

$$p_{T,\rm jet}^{\rm hadron} = \left[p_{T,\rm jet}^{\rm measured} \times f_{\rm rel} - p_T^{MI} \times (N_{\rm vtx} - 1) \right] \times f_{\rm abs}, \tag{5}$$

where $f_{\rm rel}$ is the relative jet energy correction factor as a function of jet η and p_T and $f_{\rm abs}$ is the absolute correction factor as a function of jet p_T.

At CDF, the photon-jet and Z-jet p_T balances are used to cross-check the jet energy scale in data and Monte Carlo events. When photon-jet and Z-jet balances are formed, tight cuts are made on the second jet p_T and $\Delta\phi$ between the photon/Z and jet, $p_{T,2} < 3$ GeV/c and $\Delta\phi > 3$ (rad), to suppress gluon radiation in the events that may affect the p_T balance [33]. With the cuts discussed above, the Pythia and Herwig event simulation show differences in p_T balance at the level of ∼2%–3%, a disagreement which is not well-understood yet. A better understanding of this difference will be extremely useful for future improvements of the jet energy scale uncertainty. The overall jet energy scale systematic uncertainty, evaluated as described above for central jets at CDF, is shown in Fig. 20(b); it is ∼3% at high p_T and increases at $p_T \lesssim 50$ GeV.

The hadronic decays of resonances with well-known masses, such as the W and Z bosons, can also be useful to test and to calibrate the jet energy scale. In most cases, the hadronic W and Z decays are swamped by QCD background at hadron colliders; however in, *e.g.*, $t\bar{t} \to W(\to l\nu) + \geq 4$ jets events (referred to as lepton + jets events/channel hereafter), the hadronic W resonance can be observed with only a relatively small QCD background.

Hadronic W decays in $t\bar{t}$ events were first observed in Run I in lepton + jets events in Run I [40], and in Run II these decays have been used successfully by both the CDF and DØ collaborations to finely calibrate the jet energy scale in top quark mass measurements in the lepton + jets channel [41,42]. Fig. 21 shows the di-jet mass distributions in four sub-samples of the lepton + jet event sample; the sub-samples have been divided based on the number of jets identified as b-quark jets by the standard CDF b-tagging algorithm (see Section 4.6). Events with one b-tagged jet are further divided into two classes; 1-tag(T) refers to events with four jets with $E_T > 15$ GeV and 1-tag(L) refers to events with three jets with $E_T > 15$ GeV and the fourth jet with $8 < E_T < 15$ GeV. In the CDF measurement [41], the reconstructed top mass and di-jet mass distributions are formed from $t\bar{t}$ MC events with various top mass and jet energy scales ranging from -3 to $+3\sigma$ where σ is the total jet energy scale uncertainty described above. Fits to the data without using the jet energy scale constraint from the standard procedure yield the jet energy scale $[-0.25 \pm 1.22]\sigma$, indicating that the jet energy scale from the aforementioned procedure is in good agreement with information provided by the W resonance peak in $t\bar{t}$ events.

Hadronic W/Z decays in the $W/Z + \gamma$ process have been also investigated by CDF. The observation of this signal can provide another way in which to validate and/or constrain the jet energy scale; however, the extremely large QCD background has made the observation of such a signal extremely difficult. Constraining the jet energy scale with the hadronic W resonance is a very powerful technique, and the jet energy scale uncertainty from this method will improve as more data are accumulated. However, it has to be noted that all the detailed studies presented

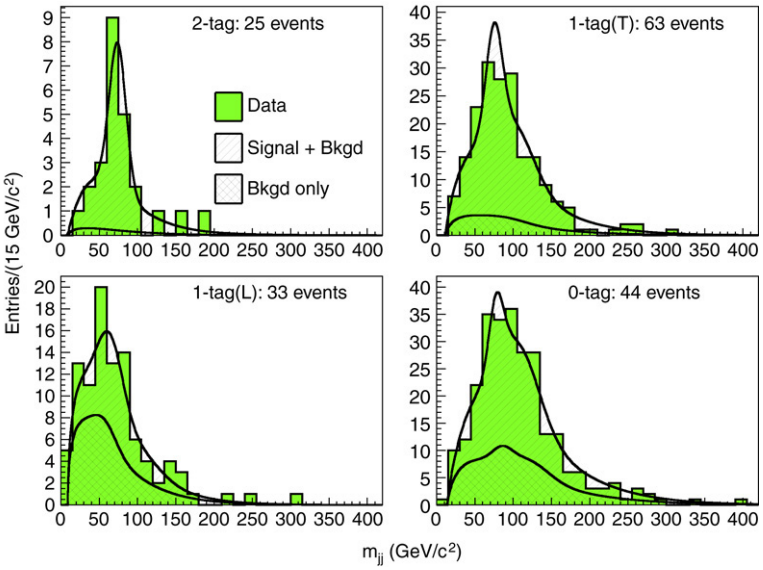


Fig. 21. Di-jet mass distributions for four subsamples in the lepton + jets sample. The signal and background shapes corresponding to the best fit of the jet energy scale cross-check are overlaid on the histograms. Clear W peaks can be observed, especially when one or more b-tags is required. Figure from Ref. [33].

above are crucial for the success of this technique, since this method relies on a good modeling of the di-jet mass distribution. Also, this technique is not able to constrain the jet energy scale over a wide range of jet p_T.

4.2.4. Summary

In this section (Section 4.2), the strategies used by the DØ and CDF collaborations to determine the jet energy scale were presented. The two collaborations used different approaches, but achieved jet energy scale determinations with similar precision. It is worth noting that, although large groups of people worked on the jet energy scale determination in both experiments, it still took more than three years to achieve the precision shown in Fig. 20 from the beginning of the Run II. High quality test-beam analyses, thorough detector simulation tuning based on the test-beam results, good planning for the jet energy calibration and validation based on the in-situ data (including the trigger implementation for the calibration/validation datasets) will be essential for the early physics analyses at the LHC.

4.3. Monte Carlo tuning

Most of the analyses at hadron colliders use parton-shower Monte Carlos in order to model the signal and background events, unfold the detector effects and to extract the physics quantities of interest. These parton-shower Monte Carlo programs have numbers of parameters that need to be tuned on the real data, *e.g.*, parameters that control the initial and final-state radiation, jet fragmentation, hadronization, and underlying event. Well-tuned Monte Carlos are essential for a precise measurement and proper comparison with theoretical predictions. The good modeling of jet fragmentation properties and of the underlying event in Monte Carlo events is crucial for corrections of jet energies from the calorimeter level to the hadron level. Also, in recent

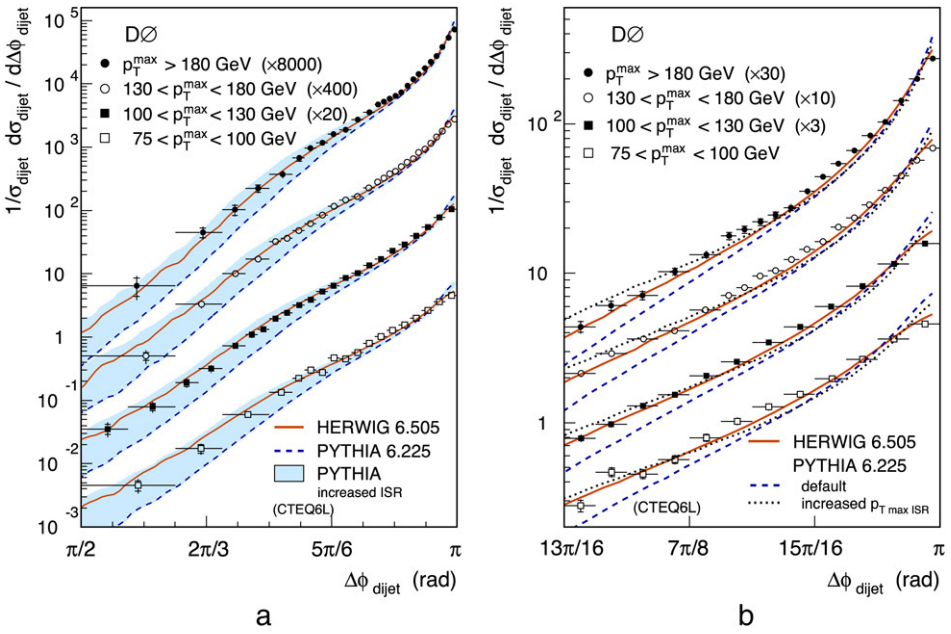


Fig. 22. $\Delta\phi$ distributions in different $p_T^{\max}$ ranges over $\Delta\phi > \pi/2$ (a) and in the peak region $\Delta\phi > 13\pi/16$. (b) Predictions from Herwig and Pythia are overlaid for comparison. Pythia predictions are shown for the PARP(67) parameter varied between 1.0 and 4.0. Figures from Refs. [6,47].

inclusive jet cross-section measurements by CDF [43–46] the effects of underlying event and hadronization are estimated using leading-order parton-shower Monte Carlo generators based on the variation of the predicted jet cross-sections after turning off the interaction between beam remnants and the hadronization in Monte Carlo events. Several measurements performed at the Tevatron, which are used for either tuning Monte Carlos or validating the tunings, are discussed below.

4.3.1. Di-jet angular decorrelations

DØ made a measurement of the azimuthal angle between the two leading jets, $\Delta\phi$, in Run II [47]. This provides an excellent testing ground for the study of multiple gluon radiation effects. Studies [6,47] have shown that such distributions are not sensitive to underlying event and hadronization effects. Near the $\Delta\phi \sim \pi$ peak the distribution is sensitive to soft radiation with small p_T, and the tail at small $\Delta\phi$ is sensitive to hard radiation with high transverse momentum. The measured $\Delta\phi$ distributions, $(1/\sigma_{jj})(\mathrm{d}\sigma_{jj}/\mathrm{d}\Delta\phi)$, for four different $p_T^{\max}$ (largest jet p_T in an event) ranges are shown in Fig. 22 along with the predictions from Herwig and Pythia. Jets are reconstructed with the midpoint cone algorithm with a cone radius of $R_{\mathrm{cone}} = 0.7$. The measured distributions have a sharp peak at $\Delta\phi \sim \pi$ and the peaks are sharper at higher jet p_T mainly due to the running of $\alpha_s(p_T)$. The default Herwig predictions give a reasonable description of the data over the whole $\Delta\phi$ range in all $p_T^{\max}$ regions; however, the default Pythia gives sharper distributions than data in all $p_T^{\max}$ regions, and Pythia provides a better description of the data when ISR is enhanced (see Fig. 22). The predictions are found to be insensitive to FSR tunes, and the measurement provides a good tool for ISR tuning. Monte Carlo predictions from Sherpa [48]

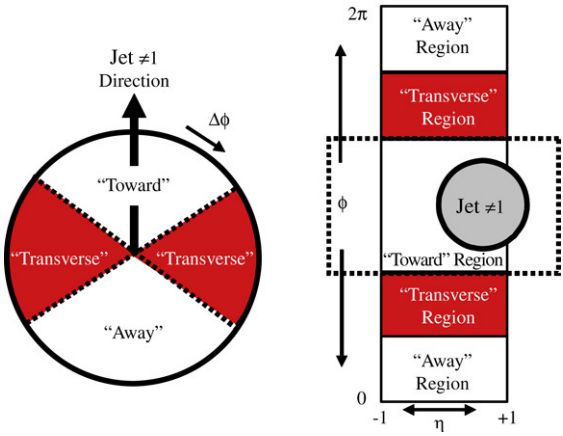


Fig. 23. Definitions of the "toward", "away" and "transverse" regions. The angle $\Delta\phi = \phi - \phi_{\text{jet\#1}}$ is the relative azimuthal angle between charged particles and the direction of the leading jet. The transverse region is defined by $60^\circ < |\Delta\phi| < 120^\circ$ and $|\eta| < 1$. Figure from Ref. [6].

and Alpgen [49,50] have also been tested against this data, as have NLO pQCD predictions from NLOJET++ [35]. They were found to give a reasonable description of the data.

4.3.2. Underlying event

Many of the important observables at a hadron collider, including jets, are sensitive to the underlying event, and so a good understanding of the underlying event is needed for precision measurements. A series of studies have been made on underlying events by CDF in Run I [2, 51] and Run II [6,52]. These studies made use of the topological structure of hadron–hadron collisions to study the underlying event; measurements were made on event activities in the "transverse" region with respect to the jet axis in jet events; this region is most sensitive to the underlying event. The geometry of one study is shown in Fig. 23, where the "transverse", "towards" and "away" regions have been defined with respect to the direction of the leading jet. The TransMAX (TransMIN) region refers to the transverse region containing the highest (lowest) scalar p_T sum of charged particles. The study made use of two classes of events, (1) "leading jet" event, in which there is no restriction on the second and third jet, and (2) "back-to-back" events, in which there are at least two jets with $p_T > 15$ GeV, the leading two jets are nearly back-to-back ($\Delta\phi(\text{jet1, 2}) > 150^\circ$) with $p_T(\text{jet2})/p_T(\text{jet1}) > 0.8$, and $p_T(\text{jet3}) < 15\ \text{GeV}/c$.

The density of transverse momentum carried by charged particles, $\mathrm{d}p_T/\mathrm{d}\eta\mathrm{d}\phi$, in both the TransMAX and TransMIN regions in the leading jet events and back-to-back events is shown in Fig. 24. In the "back-to-back" events, contributions from hard components (initial and final-state radiation) to the "transverse" region are suppressed, and the sensitivity to the underlying event is increased. In the study, the TransMAX and TransMIN regions were also used in order to better separate the hard components from the underlying event.

We expect that the MAX region will pick up the hardest ISR or FSR, and thus the MIN region will be more sensitive to the underlying event, which is indicated in Fig. 24; $\mathrm{d}p_T/\mathrm{d}\eta\mathrm{d}\phi$ in the MAX region increases with increasing leading jet p_T in leading jet events but, on the other hand, $\mathrm{d}p_T/\mathrm{d}\eta\mathrm{d}\phi$ in the MIN region stays rather flat with leading jet p_T. Please note that NLO can contribute, by definition, only to the transMAX region, and not to the transMIN. At least in the measurements at the Tevatron, the transMIN region has a level similar to that of an active

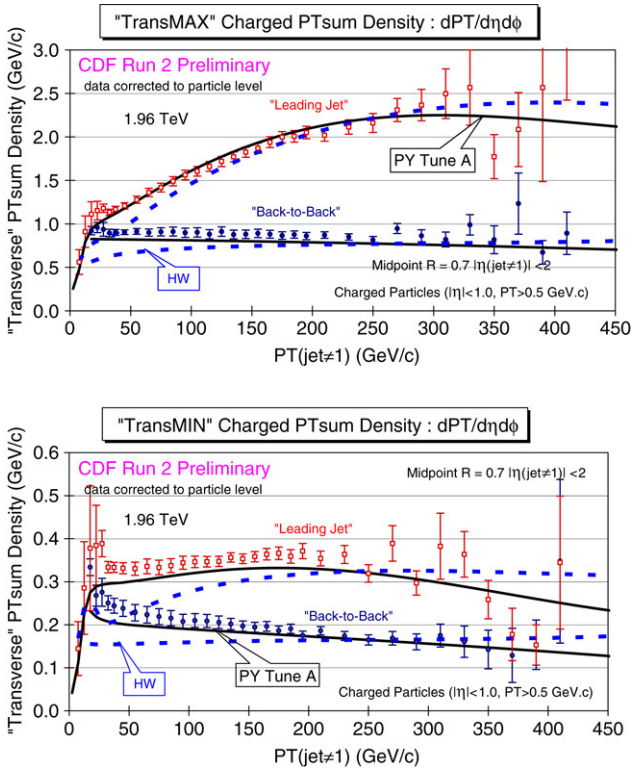


Fig. 24. The sum p_T of charged particles inside the TransMAX (top) and TransMIN (bottom) regions, as a function of the leading jet p_T, in the leading jet events and back-to-back events, in data, Pythia Tune A and Herwig (without multiple parton interactions). Figure from Ref. [52].

minimum-bias event, so the results indicate that higher-order (beyond NLO) radiation effects are relatively small. Fig. 24 shows that contributions from FSR and ISR are clearly suppressed in back-to-back events compared with the leading jet events. The MAX and MIN $\mathrm{d}p_T/\mathrm{d}\eta\mathrm{d}\phi$ are somewhat falling with increasing jet p_T at high jet p_T which could be due to a saturation of the multiple parton interactions. All these features are fairly well-described by the tuned Pythia Monte Carlo (*e.g.*, Tune A [53]).

4.3.3. Jet shapes

The jet shapes, *i.e.*, the energy flow inside jets, in inclusive jet data were studied by CDF in Run I [13] and Run II [25]. The jet shape is dominated by gluon emissions from the primary outgoing parton, and it depends on the type of the parton originating the jet, *i.e.*, quark or gluon. In hadron–hadron collisions, the jet shape is also sensitive to initial-state radiation and the underlying event. Fig. 18 in Section 3.5 showed the p_T fraction in a cone jet of radius 0.7 that lies in the annulus from 0.3 to 0.7 as a function of jet p_T; data are compared with parton-shower Monte Carlo predictions. Jets become narrower with increasing jet p_T due to several different factors: (1) power corrections that tend to broaden the jet fall as $1/p_T$ or $1/p_T^2$, (2) the fraction of jets originating from quarks increases with increasing jet p_T and (3) the probability of QCD radiation decreases as $\alpha_s(p_T)$. The measured jet shapes are well-described by the Pythia Monte Carlo with Tune A parameters, which were obtained based on the underlying event study

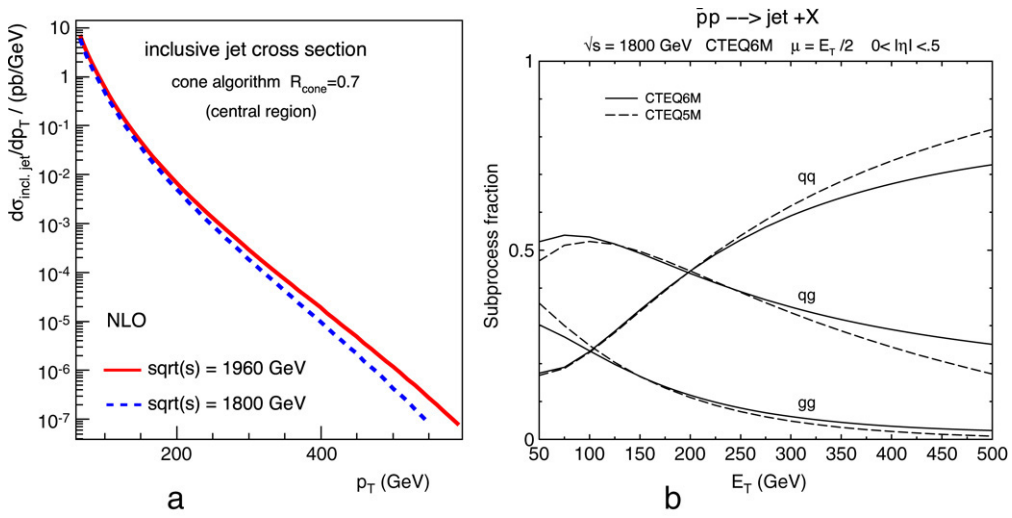


Fig. 25. (a) The inclusive jet cross-sections at $\sqrt{s} = 1.8$ and 1.96 TeV from NLO pQCD predictions using the CTEQ6.1M pdf (figure from Ref. [54]). (b) The subprocess contributions to inclusive jet production at the Tevatron for the CTEQ5M and CTEQ6M pdfs (figure from Ref. [55]).

made in the "transverse" region away from jets. It is the good description of the jet fragmentation and underlying event properties by the Monte Carlo events that allows the reliable evaluation of the unfolding correction of measurements to the hadron level and also the estimation of the hadronization and underlying event corrections.

4.3.4. Summary

In this section (Section 4.3), a few examples of the measurements that were useful for Monte Carlo tunings were presented. Most of analyses at the Tevatron experiments rely on parton-shower Monte Carlo event generators, and they benefited from these studies. At the LHC experiments, these measurements must be repeated since the scaling of these tunings to higher center-of-mass energies are not straightforward.

4.4. Inclusive jet cross-sections

The inclusive jet cross-section has been extensively studied at the Tevatron in both Run I and Run II. The differential inclusive jet cross-section at the Tevatron provides the highest momentum transfers currently attainable in accelerators. It is sensitive to a wide range of new physics, such as quark compositeness, and also tests perturbative QCD calculations over more than eight orders of magnitude in the cross-section. Due to greater statistics compared to Run I and the higher center-of-mass energy, the reach in transverse momentum has increased by approximately 150 GeV (see Fig. 25(a)). Measurements of the inclusive jet cross-section have also been shown to have a large impact on global pdf analyses [55], especially on the determination of gluon densities at high x, as the inclusive jet cross-section has sizable contributions from the $qg \to qg$ subprocess, even at high jet p_T, as shown in Fig. 25(b). In the CTEQ6M global fit [3], the full inclusive jet data sets from the Run I measurements of CDF and DØ were included, resulting in the observed enhancement of the gq subprocess compared to the predictions derived from CTEQ5M [56].

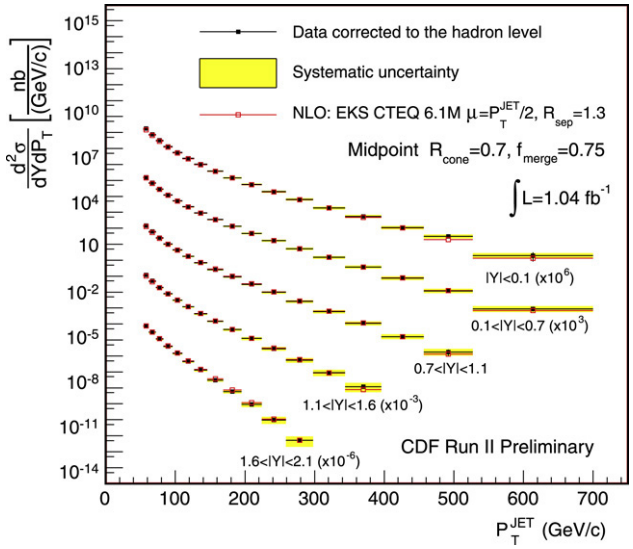


Fig. 26. The inclusive jet cross-sections measured with the midpoint algorithm by CDF in Run II [46].

The inclusive jet cross-sections measured by the CDF collaboration in Run II are shown in Fig. 26, as a function of the jet p_T in five rapidity regions. The measurement uses the midpoint cone algorithm with a cone radius of 0.7 [44,46].[5] A similar measurement was performed by DØ as well [57]. For comparisons of data to theory, the calorimeter tower energies clustered into a jet must be first corrected for the detector response and multiple $p\bar{p}$ contributions. These corrections are discussed in Section 4.2. After an additional correction that accounts for the smearing effects due to the finite energy resolution of the calorimeter (unfolding), the jet cross-sections are corrected to the hadron level. For data to be compared to a parton-level calculation, either the data must be corrected from the hadron level to the parton level or the theory must be corrected to the hadron level. In this paper, we discuss in the latter scheme; the former just involves the inverse corrections.

These corrections are intended to account for two effects: (1) "underlying event", *i.e.*, the energy not associated with the hard scattering, and (2) "hadronization", *i.e.*, a loss of energy outside a jet due to the hadronization process. In recent analyses on inclusive jet cross-sections by CDF [43–46], the hadronization corrections are evaluated by comparing the results obtained from Pythia at the hadron level to the results from Pythia when the underlying event and the parton fragmentation into hadrons has been turned off. The underlying event energy is due to the interactions of the spectator partons in the colliding hadrons and the size of the correction depends on the size of the jet cone. It is ∼1.5–2 GeV for a cone of radius 0.7 and is similar to the amount of energy observed in an arbitrarily placed cone of this size in minimum-bias events with a high track multiplicity.

The hadronization correction accounts for hadrons outside the jet cone originating from partons whose trajectories lie inside the jet cone; it does not correct for the effects of hard gluon

[5] The measurements shown here include the search cone step covered earlier in this review, as the measurements without the search cone are currently still underway. As discussed, the differences will be relatively small and will be confined mostly to lower transverse momentum.

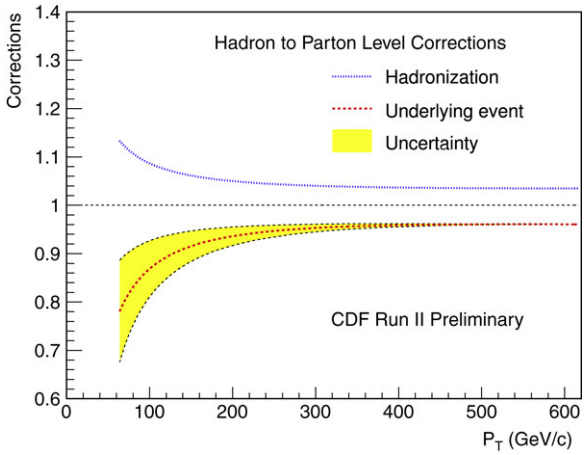


Fig. 27. Magnitude of the fragmentation and underlying event corrections used to correct the inclusive jet cross-sections measured by CDF as a function of jet p_T, for a cone size $R = 0.7$.

emission outside the jet cone, which are already accounted for in the NLO prediction.[6] The numerical value of the hadronization energy is roughly constant at 1 GeV for a cone of radius 0.7, independent of the jet transverse momentum. This constancy may seem surprising. However, as the jet transverse momentum increases, the jet becomes more collimated; the result is that the energy in the outermost annulus (the origin of the hadronization energy) is roughly constant. The evaluation of these two effects are reliable only if the Monte Carlo events provide a reasonable description of the underlying event and of the jet fragmentation and hadronization properties. Studies on underlying event and jet fragmentation are presented in Section 4.3.

The two effects (underlying event and hadronization) go in opposite directions, so there is a partial cancellation in the correction to parton level. For a jet cone of 0.7, the underlying event correction is larger, as seen in Fig. 27, for the case of inclusive jet production at CDF. For a jet cone radius of 0.4, the hadronization correction remains roughly the same size but the underlying event correction scales by the ratio of the cone areas; as a result the two effects basically cancel each other out over the full transverse momentum range at the Tevatron. Note also, as illustrated in Fig. 27, that the magnitudes of the fractional corrections are relatively insensitive to the value of p_T for jet momenta larger than 200 GeV. Although the fractional changes in the jet p_T due to the underlying event and to hadronization decrease roughly as $1/p_T$, the jet cross-section itself is becoming much steeper and hence more sensitive to changes in the p_T. These two behaviors essentially cancel each other and lead to the observed nearly constant correction factors.

A comparison of the inclusive jet cross-section measured by CDF in Run II with the midpoint cone algorithm [43,46] to NLO pQCD predictions using the EKS [10] program with the CTEQ6.1 [55] and MRST2004 [58] pdfs is shown in Fig. 28. A renormalization/factorization scale of $(p_T^{\text{jet}}/2)$ has been used in the calculation. This is the scale at which the Run I jet data were included in the global fits [3,55], so the same scale should be used for self-consistency.

[6] Such corrections for hard gluon emission are often made, however, if the comparison is to a leading-order parton calculation, such as for the reconstruction of a $t\bar{t}$ final state.

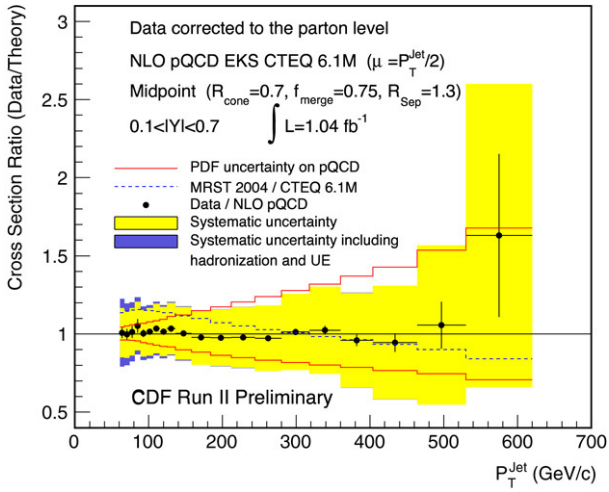


Fig. 28. The inclusive jet cross-section from CDF in Run II compared on a linear scale to NLO theoretical predictions using CTEQ6.1 and MRST2004 pdfs.

Typically, this choice of the renormalization/factorization scale leads to the highest predictions for inclusive jet cross-sections at the Tevatron. There is good agreement with the CTEQ6.1 predictions over the transverse momentum range of the prediction. Note that CTEQ6.1M gluon density is already enhanced at large x, as compared to previous pdfs, due to the influence of the Run I jet data from CDF and DØ. This enhanced gluon provides a good agreement with the high p_T Run II measurement as well which is extended approximately by 150 GeV in jet p_T. The MRST2004 pdfs also contain an enhanced higher x gluon, which leads to reasonable agreement with the measurements. The red curves indicate the pdf uncertainty for the prediction using the CTEQ6.1 pdf error set. The yellow band indicates the experimental systematic uncertainty, which is dominated by the uncertainty in the jet energy scale (on the order of 3% as shown in Fig. 20). The purple band shows the effect of the uncertainty due to the hadronization and underlying event, which is visible only for transverse momenta below 100 GeV. In Fig. 29, the jet cross-sections measured with the midpoint cone algorithm are shown for the five rapidity regions of the CDF experiment. Good agreement is observed in all rapidity regions with the CTEQ6.1 predictions. It is also important to note that for much of the kinematic range, the experimental systematic errors are less than pdf uncertainties; thus, the use of this data in future global pdf fits should serve to further constrain the gluon pdf.

While this measurement has been carried out, a new seedless cone algorithm, SISCone [16] (see also Section 3.1.3), has become available. The differences in the inclusive jet cross-sections between the midpoint algorithm[7] and SISCone algorithm are evaluated in the Pythia Monte Carlo samples both at the hadron level and parton level and are shown in Figs. 30 and 31. The hadron-level inclusive jet cross-sections are obtained from the Pythia Monte Carlo samples generated with the Tune A parameters, and the parton-level inclusive jet cross-sections are obtained from

[7] For comparisons with the SISCone algorithm, the midpoint algorithm without the search cone step is used in order to investigate the effects of seeds and slight differences in the merging procedure of the overlapping stable cones only. Please also note that DØ uses the midpoint algorithm without the search cone step, and a measurement using the midpoint algorithm without the search cone step is being finalized at CDF too.

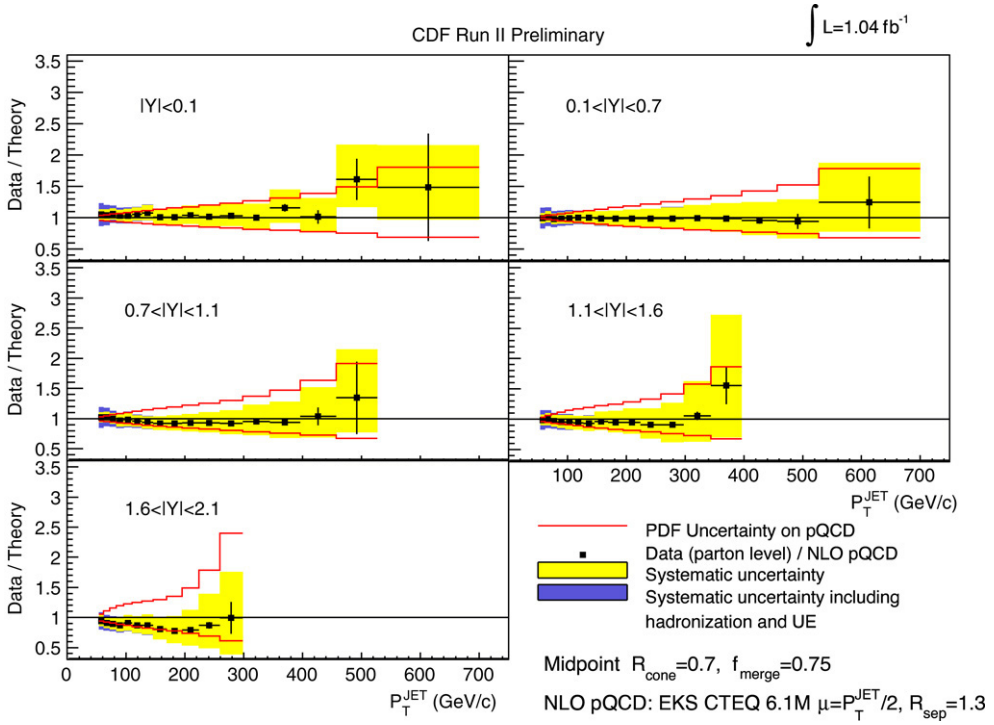


Fig. 29. The inclusive jet cross-section from CDF in Run II, for several rapidity intervals using the midpoint cone algorithm, compared on a linear scale to NLO pQCD predictions using CTEQ6.1 pdfs.

the Pythia Monte Carlo samples generated with the underlying event turned off. The cross-sections at these levels are used to determine the hadronization corrections. Fig. 30 indicates that the hadron-level inclusive jet cross-sections are different between the SISCone and midpoint algorithms by 5% at low jet p_T with the differences decreasing with increasing jet p_T; in the meantime, the differences are less than 1% at the parton level for any jet p_T (see Fig. 31). This indicates that the hadronization correction is different between the SISCone and midpoint algorithms by up to 5%; however, the impact on comparisons between the measurement and NLO predictions is negligible (less than 1%).

CDF has also made measurements of the inclusive jet cross-sections with the k_T algorithms in Run II. In Fig. 32, the experimental jet cross-sections using the k_T algorithm from CDF Run II [44,45] are compared to NLO pQCD predictions using the Jetrad [34] program. Similar to the measurement using the midpoint cone algorithm, good agreement is also observed between data and theory. A comparison of these measurements in the central region is presented in Fig. 33 [46]. In order to require that the algorithms use approximately the same size in $y - \phi$ space, the cone size for the midpoint algorithm and the D parameter for the k_T algorithm will both be taken as 0.7 ($R_{\text{cone}} = D = 0.7$). It is important to note that we expect different predictions on the cross-sections for jets clustered with the midpoint and k_T algorithms when the parameters R_{cone} and D are set equal. Fig. 33 shows the ratio of the measured inclusive jet cross-sections for jets clustered with the k_T algorithm to the result for jets clustered with the midpoint algorithm. Only statistical errors are shown assuming no correlation between the two measurements. The prediction of this ratio from NLO pQCD calculations using fastNLO [59] is also shown, and good agreement is

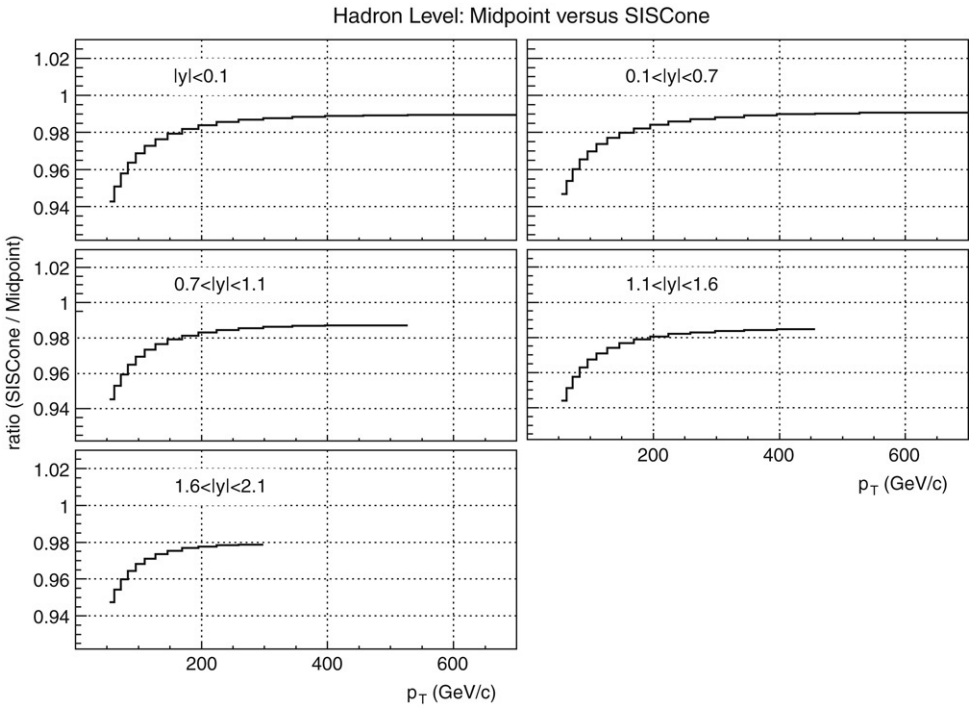


Fig. 30. The ratio of the hadron-level inclusive jet cross-section with the SISCone algorithm to that with the midpoint algorithm in five rapidity regions in Pythia Monte Carlo samples generated with the Tune A parameters.

observed. As expected, for $D = R_{\text{cone}}$, the k_T cross-section, at the parton level, is slightly smaller than the cross-section using the cone algorithm.[8]

Fig. 34 shows the ratio of the hadron to parton-level correction derived with the k_T algorithm to the one derived with the midpoint algorithm. These corrections were derived from Pythia Tune A [53] (see Section 4.3). The multiplicative corrections are both smaller than one, so the observed ratio indicates that the size of the correction obtained with the k_T algorithm is larger (*i.e.*, farther away from unity) than the correction obtained with the midpoint algorithm. The consistency of the data–theory comparisons between the k_T inclusive jet cross-section measurement and the midpoint measurement indicate the robustness of the obtained results and adds credence to the fact that the jet definitions are made consistently at the parton and detector levels.

4.5. W/Z+jets

The production of a W or Z boson in conjunction with jets is an interesting process in its own right as well as a background to many Standard Model (SM) and non-SM physics signals. Jet multiplicities of up to 7 have been measured at the Tevatron. Production of $W/Z+$ jets at the Tevatron is dominated by gq initial states. The NLO cross-sections have been calculated

[8] For the NLO pQCD prediction for jets clustered with the midpoint algorithm, the R_{sep} parameter of 1.3 is used. If the R_{sep} parameter is set to 2, *i.e.*, R_{sep} is not used, the result is a larger difference between the two algorithms at NLO (*i.e.*, setting $R_{\text{sep}} = 2$ results in a larger jet cross-section as seen in Fig. 8).

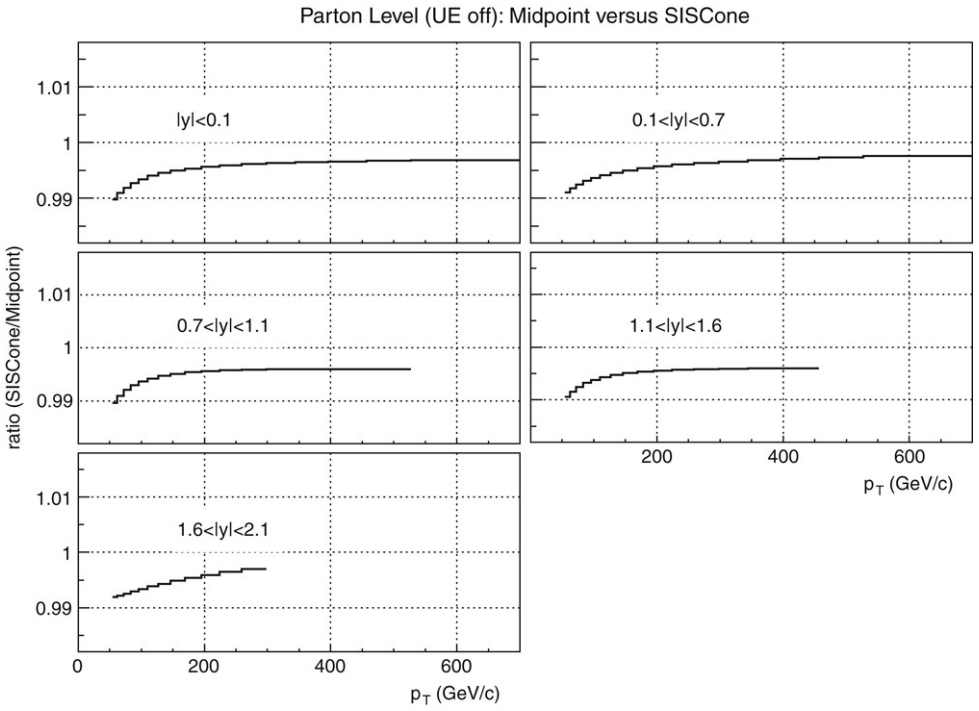


Fig. 31. The ratio of the parton-level inclusive jet cross-section with the SISCone algorithm to that with the midpoint algorithm in five rapidity regions in Pythia Monte Carlo samples in which the underlying event has been turned off.

only for W/Z+up to 2 jets; predictions for the higher jet multiplicity final states are accessible through matrix element (+ parton-shower) predictions and in fact can be considered as a prime testing ground for the accuracy of such predictions as well as for measurements of α_s. The jet multiplicity distribution for $W + n$ jets measured at the Tevatron is shown in Fig. 35. The measurements (and predictions) were performed with a jet cone of radius 0.4 and a minimum transverse energy requirement of 15 GeV. A smaller jet cone size is preferred for final states that may be "complicated" by the presence of a large number of jets. Also shown in the figure are the NLO predictions (for jet multiplicities less than 3), parton-shower Monte Carlo predictions and LO matrix element + parton-shower Monte Carlo predictions. The NLO predictions are able to describe the absolute rates for jet production (for up to 2 jets) while the LO matrix element + parton-shower Monte Carlo predictions can describe the relative jet multiplicity rates.

A recent measurement of $W \to e\nu + \geq n$ jets from CDF is shown in Fig. 36. In this analysis, the data have been reconstructed using the JETCLU cone algorithm with a cone radius of 0.4. The data have been compared, at the hadron level, to predictions using matrix element information from Alpgen [49,50], and parton-shower and hadronization information from Pythia [4]. The agreement is reasonable, although the data has a tendency to be somewhat softer than the predictions. This data has been corrected to the hadron level; this makes it convenient for comparison to any hadron-level Monte Carlo prediction.[9] Data corrections to the hadron level

[9] As mentioned before, the corrections for underlying event and for fragmentation basically cancel each other out for a cone of radius 0.4, for inclusive jet production, so that the hadron-level predictions are essentially parton-level predictions as well. It was found that the same statement is essentially true also for $W+$ jet production.

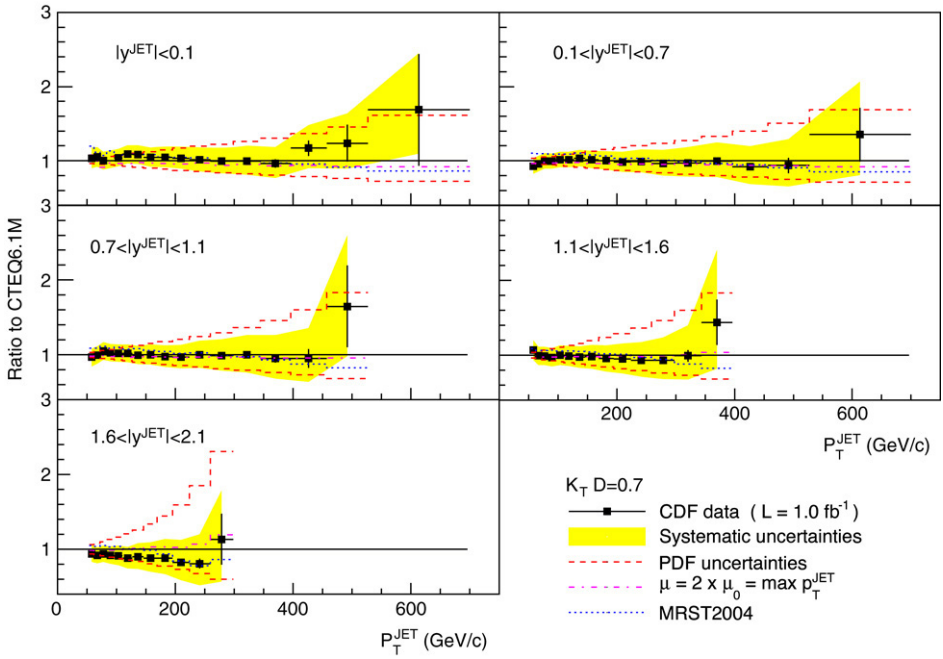


Fig. 32. The inclusive jet cross-section from CDF in Run II, for several rapidity intervals using the k_T jet algorithm, compared on a linear scale to NLO theoretical predictions using CTEQ6.1 pdfs. Figure is from Ref. [45].

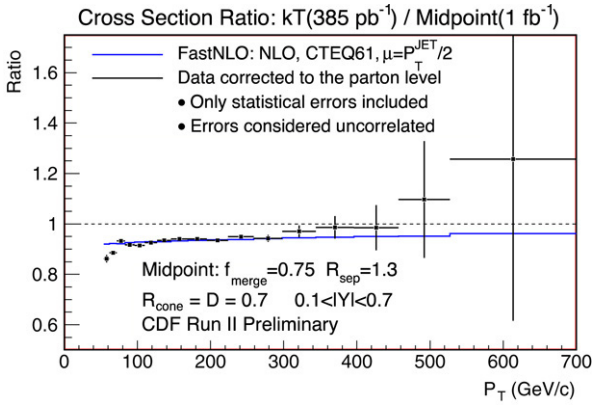


Fig. 33. Ratio of the inclusive jet cross-section measured with the k_T algorithm to that measured by the midpoint algorithm (black points). The prediction of this ratio from the NLO program fastNLO is also shown in the figure (blue line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(and/or parton level) should be the norm for measurements at the Tevatron and LHC, in order for the best interplay between theory and experiment.

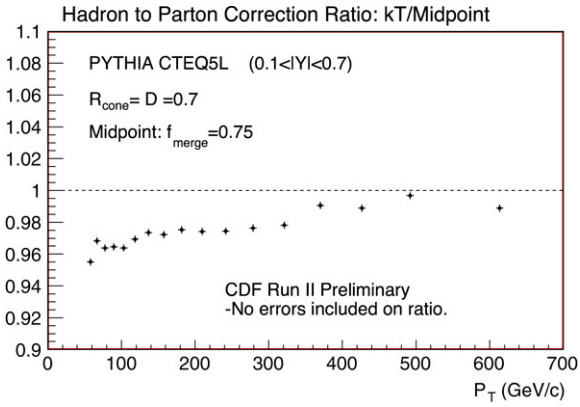


Fig. 34. Ratio of the hadron to parton-level correction derived with the k_T algorithm to that derived with the midpoint algorithm. The multiplicative corrections are both smaller than one. The correction derived with the k_T algorithm is larger (farther away from unity). The corrections were derived from Pythia Tune A.

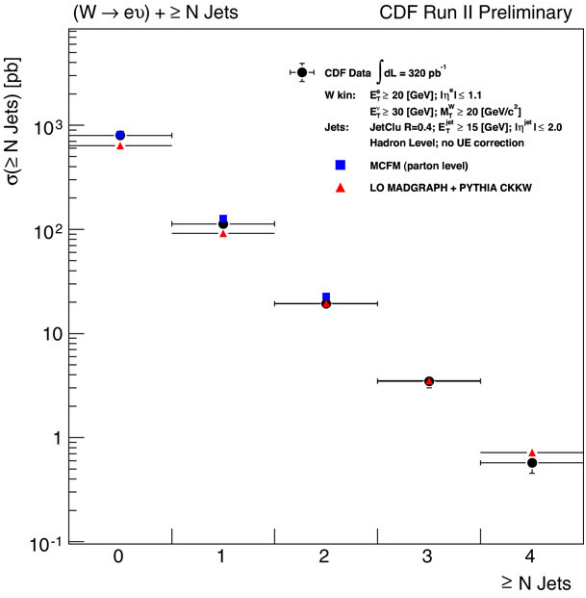


Fig. 35. The rate of production of $W+n$ jets at CDF. The measurements and predictions were performed with a cone jet of radius 0.4 and with a requirement of 15 GeV/c or greater. The MCFM [60] predictions are absolutely normalized. The CKKW [61] predictions are normalized to the first bin. A scale of 10 GeV/c has been used for the matrix element/parton-shower matching.

4.6. *Heavy flavor jets*

Many of the interesting final states at the Tevatron, such as $t\bar{t}$ decays and $H \to b\bar{b}$, involve b-quark jets (b-jets). $W/Z+b$ and $\gamma+b$ processes are also interesting; they are major backgrounds in Standard Model Higgs or SUSY searches, and they are also sensitive to the b content of

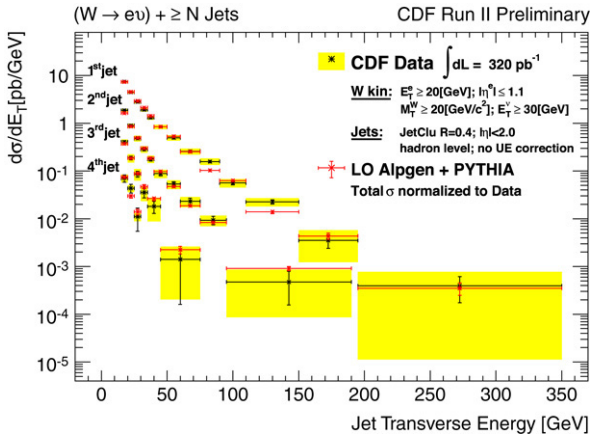


Fig. 36. A comparison of the measured cross-sections for $W+ \geq n$ jets in CDF Run II to predictions from Alpgen + Pythia. The experimental cross-sections have been corrected to the hadron level.

protons. There are two main challenges in the analyses that deal with b-jets; (1) the b-jet identification (b-tagging) and (2) the energy measurement of the b-jet.

There are characteristics of b-jets that differentiate them from light flavor and charm jets:

- the long lifetime of the b quark
- the large mass of B hadrons
- the energetic semileptonic decay of B hadrons.

The algorithms that identify b-jets exploit these characteristics to separate b-jets from the other jets. The most widely used algorithm to tag heavy flavor jets at CDF is the secondary vertex algorithm, often referred to as SecVtx [62,63]. Because of their long lifetime, b quarks typically decay a measurable distance from the primary interaction point, and so the algorithm reconstructs the decay vertices (secondary vertices) using a minimum of two or three tracks with an impact parameter significance ($d_0/\sigma(d_0)$) greater than 3.0 or 2.0, respectively. The impact parameter (d_0) is the minimum distance between the track and the primary vertex in the plane transverse to the beam direction and $\sigma(d_0)$ is its uncertainty. The 2-dimensional displacement of the secondary vertex from the primary interaction point projected along the jet axis (L_{2D}) is then measured; a jet is b tagged if the vertex has L_{2D} significance larger than 7.5, where the uncertainty on L_{2D} includes contributions from both the primary and secondary vertex fits [63]. When considering a b-tagging algorithm, it is important to understand how often one tags a b-jet in the data (*i.e.*, tagging efficiency) and how many of the tagged jets are actually from non-b's (*i.e.*, mis-tagging rate). These features for SecVtx at two operating points are shown in Fig. 37.

There are other b-tagging algorithms developed at CDF. The algorithm called JetProbability considers d_0 of each track within a jet and constructs a probability that a given jet is consistent with coming from a zero-lifetime source [64]. The soft lepton tagging algorithm identifies b-jets by looking for a semileptonic B hadron decay within a jet [65]. Efforts to combine these tagging tools using a multivariate technique like a neural network are also underway.

Since b-jets have quite different characteristics from other jets (light quark or gluon), special care is necessary for the energy measurement. The energy correction from calorimeter jets to hadron-level jets or to the parent parton is different for b-jets than for generic jets because of different parton shower and fragmentation properties, and also due to the presence of

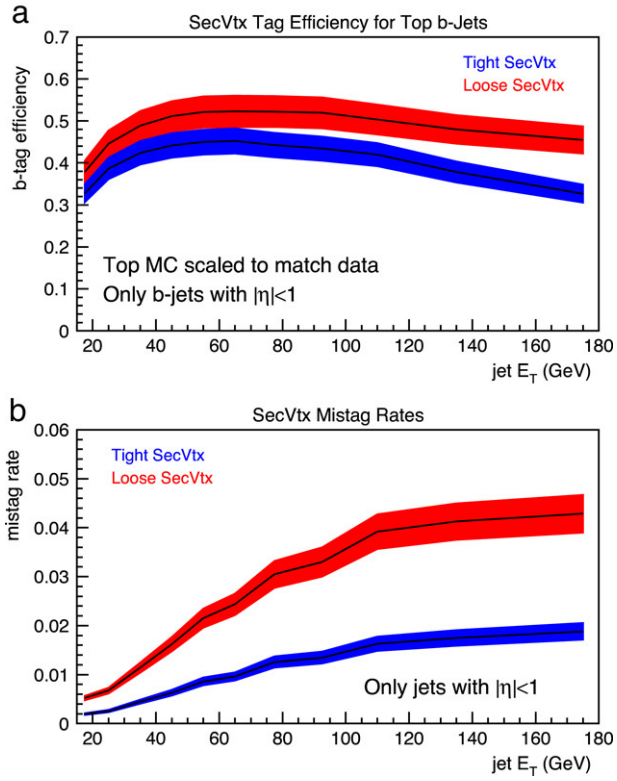


Fig. 37. (a) b-tagging efficiency for b-jets in top decays and (b) mis-tagging rate as a function of jet E_T at CDF (figures from Ref. [63]).

semileptonic decays. Both CDF and DØ generally rely on MC simulation to model the b-jet energy scale. In the CDF top quark mass measurements [41], additional uncertainties are evaluated for the b-jet energy scale: (1) uncertainties in energy response arising from uncertainty in the B meson semileptonic branching ratios, (2) uncertainties arising from the imperfect knowledge of the fragmentation properties of b-quarks, and (3) uncertainties arising from the different color flow associated with b-jets produced in top quark decay. The b-jet energy scale uncertainties from these sources are evaluated by changing the relevant parameters in the MC based on the constraints from other experiments, and they yield an additional 0.6% uncertainty in total.

Possible ways to test the b-jet energy scale in $p\bar{p}$ data would be to use the photon-b-jet p_T balance or the $Z \to b\bar{b}$ resonance. DØ recently made a preliminary measurement on b-jet energy scale in γ+b-jet events using the missing E_T projection fraction (MPF) method [66]. The MPF method is described in detail in Section 4.2.2. The study suggested that b-jets need additional energy corrections of as much as 10% at energies around 20 GeV and about 5% at energies of 150 GeV.

CDF and DØ have also extracted the $Z \to b\bar{b}$ signal and are seeking to use it to test and calibrate the b-jet energy scale [67,68]. At CDF, a dedicated trigger was implemented to collect a large sample of Z decays to b-quark pairs by requiring two tracks with displaced vertices and two jets. Di-jet events with both jets being tagged as b-jets were selected offline.

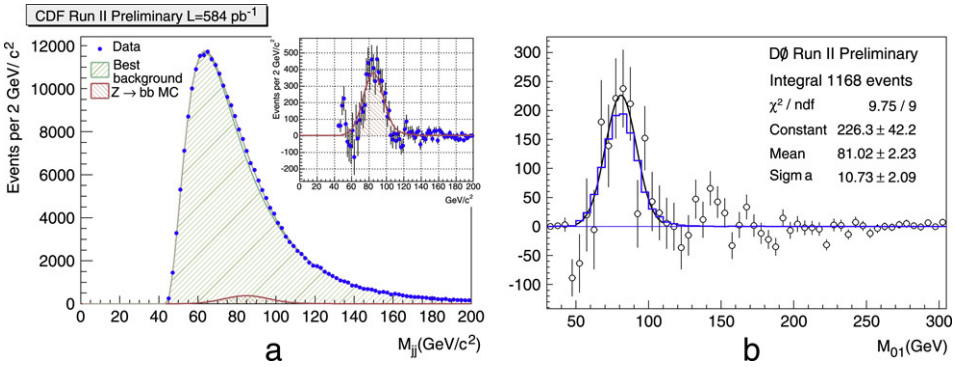


Fig. 38. (a) the $b\bar{b}$ di-jet mass distribution measured in CDF data (points) and the estimated QCD $b\bar{b}$ background (green) and $Z \to b\bar{b}$ signal shape (red). The inset shows data minus background distributions compared to the $Z \to b\bar{b}$ signal shape. (b) $Z \to b\bar{b}$ peak extracted from DØ data (points), compared to the shape of $Z \to b\bar{b}$ distribution in MC (histogram). Figures from Refs. [67,68]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The signal was extracted by fitting the data with the QCD background shape computed using untagged data passing the same kinematic selection together with the $Z \to b\bar{b}$ signal and the $Z \to b\bar{b}$ signal shape computed with Pythia. The extracted $Z \to b\bar{b}$ signals by CDF and DØ are shown in Fig. 38. The measured data/MC scale factor for b-jet energy scale for CDF is $k = 0.974 + 0.020 - 0.018(\text{stat} \oplus \text{syst})$. The achieved uncertainty of $\sim$2% is smaller than the convolution of the generic jet energy scale and additional b-jet specific uncertainties of $\sim$4 %, and thus the $Z \to b\bar{b}$ signal can provide a good constraint on the b-jet energy scale. The signal is also expected to serve as a tool to test b-tagging algorithms and also for improvements in the jet energy measurement algorithms which will be discussed in Section 4.7.

4.7. Particle flow type approaches

In the Tevatron experiments, by default, jets are reconstructed by running a jet clustering algorithm on energy depositions in the calorimeters, *i.e.* basically only calorimeter information is used in the standard jet reconstruction and energy measurement. The H1 collaboration at HERA [69] and OPAL collaboration at LEP [70], among others, have successfully used an algorithm that incorporates tracks in jet reconstruction, and such an algorithm has been tested at CDF and DØ as well. As discussed, the energy measurement of hadronic particles by calorimeters suffers from large fluctuations and nonlinearities; tracking information can be used to reduce such effects and to improve the energy resolution in jet energy measurements. There is also an attempt in CDF to further improve the jet energy measurement by using the information from the shower-max detector (a wire chamber placed near the shower maximum position in the central electromagnetic calorimeter) to sort out the overlapping particles like π^0 with $\pi^\pm$ [71]. Although such methods are not yet mature enough to be used in any physics analysis, they could be helpful in achieving our goal of percent precision and should be pursued further.

5. Jets at the LHC

The experience gained at the Tevatron is extremely useful in the preparation for physics analysis with jets at the LHC. However, hard scattering at the LHC is not just “rescaled”

Table 1
Expected inclusive jet production rate at LHC in one run year with an integrated luminosity $\int \mathcal{L} dt = 10\ \text{fb}^{-1}$

$p_T^{\min}$ (TeV/c)	σ (nb)	Events/year
0.2	100	$\approx 10^9$
1.0	0.1	$\approx 10^6$
2.0	1.0×10^{-4}	$\approx 10^3$
3.0	1.3×10^{-6}	≈ 10

scattering from the Tevatron. Many of the interesting physics signatures will take place with relatively low x partons and thus there will be a dominance of gluon and sea quark interactions, as compared to interactions involving valence quarks. In addition, as the initial-state partons are at low x, there is enormous phase space for gluon emission, and so a large probability for additional jets from initial-state radiation. The underlying event is also expected to be enhanced compared to the activity observed at the Tevatron, largely through an increased rate of semihard multiple-parton interactions. These interactions will contribute to the energy measured in jets produced in the hard scatter, and may often lead to the production of extra low transverse momentum jets in their own right. Thus, the LHC will be a very "jetty" environment and accurate measurements of the dynamics of the hard scattering may be challenging [1]. There is then a need for tools even more powerful than the ones used at the Tevatron to reconstruct jets. The most interesting tools focus on the reconstruction of jet shapes, thus exploiting the significantly finer readout granularity of the LHC detectors. In this section we summarize some of the related aspects presently under study, and show some expectations for the jet final state at LHC and its representation in the detectors.

5.1. Expectations for jet final states

As already mentioned, jets will be generated in basically all final states expected in pp collisions at $\sqrt{s} = 14$ TeV at the LHC. For most channels they are going to be the dominant part of the detectable signal, thus providing major input to the reconstruction of the event kinematics. Precision requirements on the jet energy scale are high compared with the Tevatron, with systematic uncertainties of less than 1% absolute needed/expected for jets reconstructed in $t\bar{t}$ production, or for jets generated at the end of long decay chains in certain SUSY models.

The kinematic reach of jets produced in QCD $2 \to 2$ processes at the LHC greatly extends that possible at the Tevatron; for example, compare Fig. 25 in Section 4.4 with the predictions for the LHC shown in Fig. 39. Even in the initial lower luminosity running phase, the jet production rate is also very high, as is summarized in Table 1. It is interesting to note that the pdf uncertainty for jets at the LHC at the highest attainable transverse momentum is similar to that for the Tevatron at its highest attainable transverse momentum.

The partonic phase space available in pp collisions at the LHC allows for a large amount of extra radiation, as can be observed in the number of jets calculated as a function of the leading jet p_T in QCD $2 \to 2$ processes (see Fig. 40). For low p_T the number of additional jets is suppressed by the p_T cut applied to jets being nonnegligible compared to the transverse momentum of the hard scattering, while the drop towards higher p_T indicates that radiation is suppressed due to the increasing dominance of qq in higher partonic x scatterings, and the subsequent lower color factors in the collision. There is also an extra suppression due to the higher x values of the incident partons.

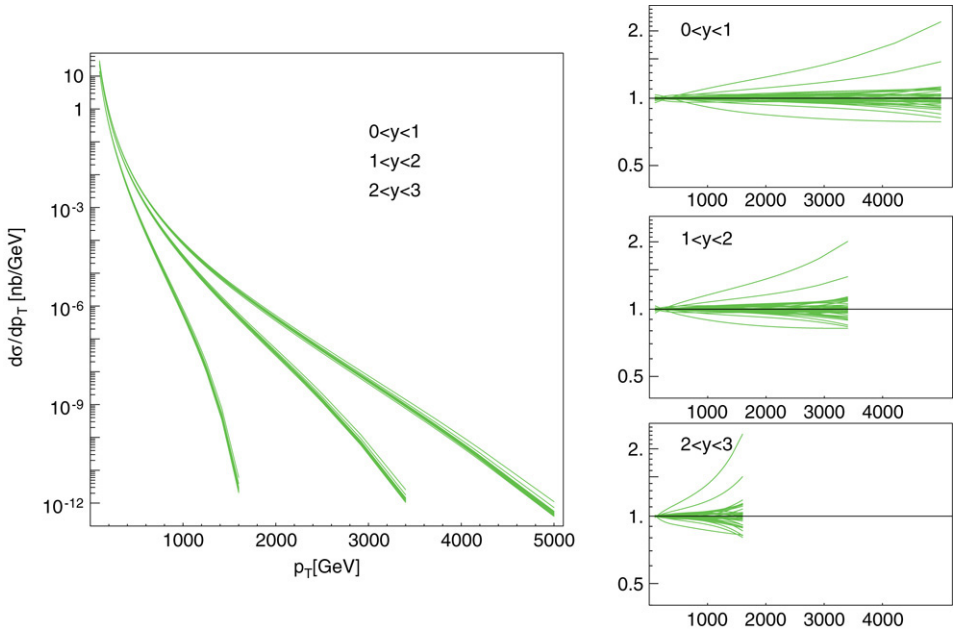


Fig. 39. NLO Inclusive jet cross-section predictions for the LHC using the EKS [10] program and the CTEQ6.1 central pdf and the 40 error pdf's (left); the ratios of the jet cross-section predictions for the LHC using the CTEQ6.1 error pdf's to the prediction using the central pdf (right).

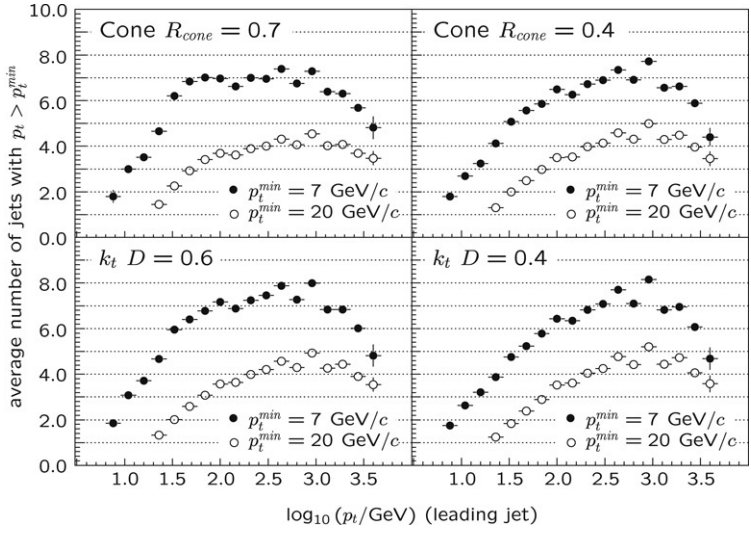


Fig. 40. Number of hadron jets in simulated QCD $2 \to 2$ LHC events, as a function of the leading jet p_T, for various jet finders. The shapes of the curves can be understood as a convolution of the p_T cut applied in jet finding, the increasing probability for extra hard gluon emission with the increasing hardness of the $2 \to 2$ scattering, the change from gluon to quark jets with increasing p_T, and finally the restricted phase space for extra jets accompanying very energetic primary jets.

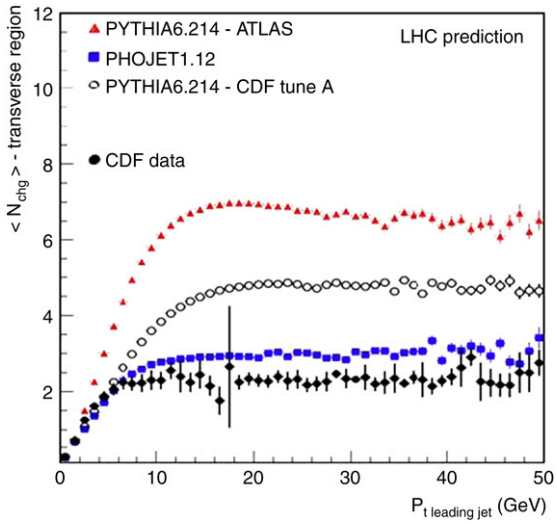


Fig. 41. Number of charged tracks in the transverse region of the QCD $2 \to 2$ interaction plane (as defined in Fig. 23) as a function of the leading jet p_T. Data from CDF[2] are shown together with model predictions for the LHC (figure from [73]).

5.2. *Jet physics environment at the LHC*

Jet measurements are affected by the presence of multiple semihard interactions from other parton–parton pairs from the proton–proton collision of interest, *i.e.* the underlying event (UE) of the same collision. In addition, the physics environment at the LHC is also affected by additional minimum-bias (MB) collisions of other proton pairs in the same bunch crossing. Both effects limit the efficiency for reconstructing the hadron-level jets (and ultimately the parton-level jets) from the hard scattering (signal) and also add complex features to the already nontrivial detector jet signals. They are also major sources for uncertainties in the present simulation-based performance estimates for LHC physics.

5.2.1. *The underlying event at the LHC*

There is a great deal of uncertainty in the level of underlying event activity expected for pp collisions at $\sqrt{s} = 14$ TeV, as can be observed in Fig. 41. This uncertainty is a major factor in estimating the quality of the reconstruction of the jet signals. Estimates derived from a recent tuning of Pythia [4] ("Pythia 6.214 CDF Tune A"), discussed previously in this paper, actually predict significantly fewer tracks from the UE at the LHC than the parameters previously used in the same Pythia version by ATLAS [73] ("Pythia 6.214 ATLAS" in Fig. 41). The determination of the level of this underlying event activity will be one of the first measurements to take place upon startup of the LHC.

An ATLAS study similar to the one carried out by CDF (and shown in Fig. 24) indicates that the sum of the transverse momentum of the charged particles in the transverse region (see Fig. 23) will vary from approximately 10 GeV/c for low jet transverse momentum to over 30 GeV/c for jet transverse momenta larger than 1 TeV/c [73].

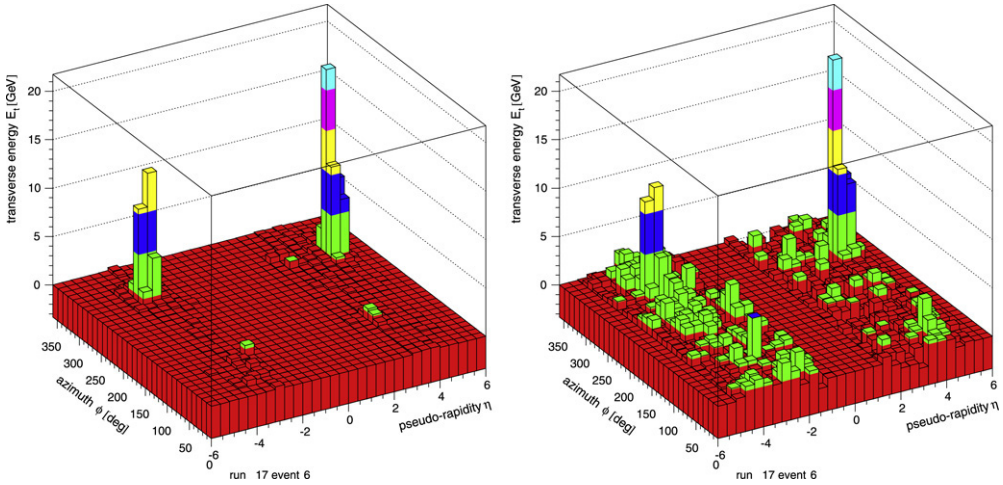


Fig. 42. Simulated forward going quark jets from Higgs production via WW scattering, without (left) and with pile-up (right) in the ATLAS forward calorimeter. The signals with $p_T < 0$ are due to the specific choice of bi-polar signal shaping functions for this detector. The signals in the central detector regions are omitted here for clarity.

5.2.2. *Minimum-bias events and pile-up*

Contrary to the underlying event energy, the additional minimum-bias interactions present at higher luminosities provide a source for p_T flow in the LHC collisions which is not correlated at all with the (triggered) hard scattering of the signal event. Its total contribution is dependent on the instantaneous luminosity. Assuming a total pp cross-section of about 75 mb and excluding single diffractive and double diffractive collisions, one can expect an average number of MB collisions of ≈23 (Poisson-distributed) at the design luminosity ($L = 10^{34}$ cm^{-2} s^{-1}), ≈4.6 at the "initial" luminosity $L = 2.0 \times 10^{33}$ cm^{-2} s^{-1}, and ≈0.02 at the LHC startup[10] ($L = 10^{31}$ cm^{-2} s^{-1}, the most recent expectation). This activity is therefore likely negligible at the LHC startup, but can produce a large number of nonsignal tracks/energy at the design luminosity.

The actual detailed effects of the minimum-bias events on the jet signal depend strongly on the calorimeter technology and readout electronics. For example, the rather slow signal formation in the ATLAS liquid argon calorimeters with typical charge collection times of about 500 ns (which has to be compared to the LHC bunch crossing time (25 ns)) and the rather large pp cross-section leads to a history of previous collision signals still visible in the actual event. The effect averages to 0 energy, due to the specifically chosen bi-polar (canceling area) signal shaping, but the "out-of-time" pile-up adds to the fluctuations. Fig. 42 shows the simulated response to two rather low p_T jets from vector boson fusion (VBF) Higgs production in the ATLAS forward calorimeters with and without full LHC luminosity pile-up added. Both jets are well-visible above the noise, but their shapes, as well as their signal amplitudes, are changed, which makes calibration in this particularly hostile region very challenging. A more quantitative estimate for the signal fluctuations introduced by pile-up in jet cones in shown in Fig. 43 [74].

[10] These estimates assume that the bunch crossing time is 25 ns, with about 3000 bunches in LHC. Less frequent bunch crossings, and a smaller number of (longer) bunches at the same stored current, as recently discussed for initial LHC running, can increase the pile-up significantly, even at lower luminosities.

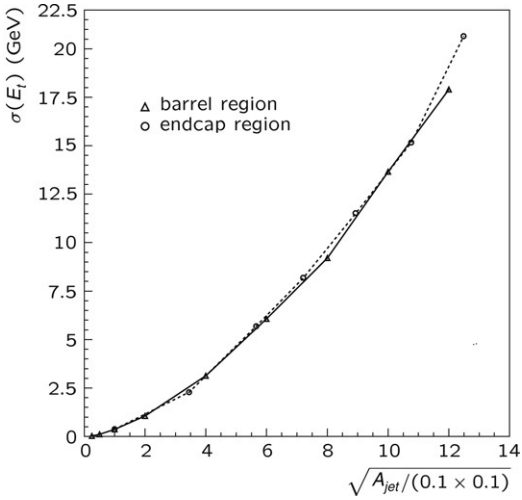


Fig. 43. Estimated signal fluctuations from full pile-up simulations, calculated as RMS in p_T for the central and endcap region in ATLAS, as a function of the square root of the jet cone area. For a typical cone size $R_{\text{cone}} = 0.7$ the fluctuation are about 12 GeV in both detector regions (taken from [74]).

5.3. Experimental aspects of jet reconstruction at the LHC

The large calorimeter systems in ATLAS and CMS at the LHC are the basic detectors for jet reconstruction. Both systems provide hermetic coverage up to pseudorapidities of ∼5. Cell sizes and readout granularity vary widely within each of the systems and introduce different limitations on the calorimeter signals used for jet finding and reconstruction. In general, though, the cell sizes are smaller than in the Tevatron calorimeters, allowing for the development of more powerful jet clustering software. Typical depths of the calorimeter systems exceed 8–10 absorption lengths. Fig. 44 gives an overview on these detectors, with some of the relevant details described below.

5.3.1. Brief look at the ATLAS and CMS calorimeters

CMS features a highly granular central electromagnetic lead-tungstate ($PbWO_4$) crystal calorimeter with very small lateral cell sizes ($\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$ within $|\eta| < 1.479$). The calorimeter consists of 61200 individual crystals and has a total depth of 25.6 radiation lengths (X_0). The crystals point back to the vertex, within a very small tilt. The electromagnetic endcap calorimeters also use $PbWO_4$ crystals, but with a coarser granularity and a rectangular pattern of fixed-sized crystals. The depth of these crystals is about $24.7X_0$. Each endcap covers the region $1.479 < |\eta| < 3.0$.

Hadronic calorimetry in CMS features tiled scintillator read-out with brass absorbers. The central hadronic calorimeter is arranged outside the central electromagnetic calorimeter, but still inside the solenoid magnet. It has a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in one depth segment within $|\eta| < 1.4$. The hadronic outer detector, a scintillator layer attached to the outside of the solenoid magnet, provides additional depth coverage for hadrons. The hadronic endcap calorimeters feature the same readout technology with decreasing granularity in η, starting with $\Delta\eta \times \Delta\phi = 0.087 \times 5^\circ$ at $|\eta| = 1.3$ to $\Delta\eta \times \Delta\phi = 0.35 \times 10^\circ$ at $|\eta| = 3.0$. The gap to the beam pipe is closed by a steel/quartz-fiber forward calorimeter with typical granularity of $\Delta\eta \times \Delta\phi \approx 0.175 \times 10^\circ$. All CMS calorimeters are noncompensating and require specific calibrations for hadrons and jets. More details on the CMS detector can be found in [75].

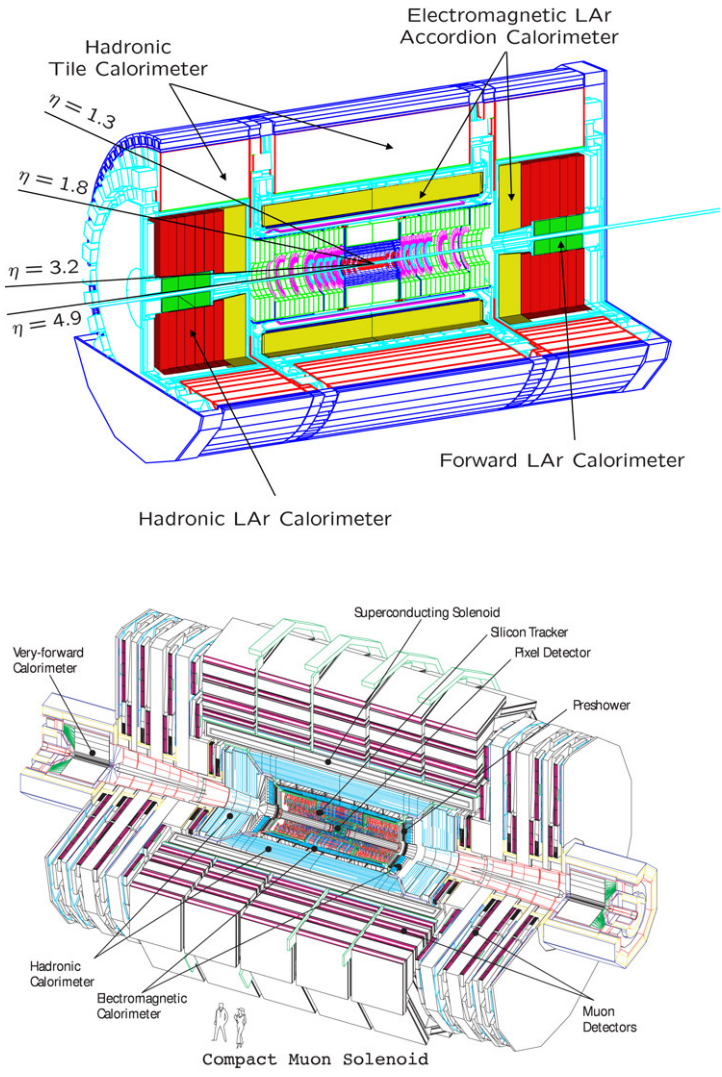


Fig. 44. The ATLAS calorimeter system (top) and CMS detector system (bottom).

The ATLAS calorimeter system features a central highly granular electromagnetic liquid argon/lead accordion calorimeters with a pointing readout geometry. Cell sizes vary from $\Delta\eta \times \Delta\phi = 0.003 \times 0.1$ in the first depth segment to $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ in the second, and $\Delta\eta \times \Delta\phi = 0.5 \times 0.025$ in the third depth sampling. Depending on η, the depth varies from 26 to 36 X_0. The central electromagnetic calorimeter is, contrary to the setup in CMS, located outside the solenoid magnet. The calorimeter covers a region of $|\eta| < 1.475$.

The electromagnetic endcap calorimeter in ATLAS features a liquid argon/lead accordion-type calorimeter with the absorber folded like a Spanish fan. It covers $1.375 < |\eta| < 3.2$, with three depth samplings up to $|\eta| = 2.5$, and two in $2.5 < |\eta| < 3.2$. The lateral size of the pointing cells is $\Delta\eta \times \Delta\phi = (0.003 - 0.006) \times 0.1$ in the first sampling, $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ in the

second, and $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$ in the third sampling, all up to $|\eta| = 2.5$. Beyond that, the two remaining samplings of the electromagnetic endcap calorimeter have $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$.

Hadronic calorimetry in ATLAS is provided by the steel/scintillator tile calorimeter in $|\eta| < 1.7$, which has three samplings with quasi-projective cells of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two, and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ in the last depth sampling. The endcap hadronic calorimeter is a liquid argon/copper parallel plate calorimeter with four depth samplings and quasi-projective cells with $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in $1.5 < |\eta| < 2.5$, and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ in $2, 5 < |\eta| < 3.2$. The net result is that the ATLAS calorimeters have six or seven depth samplings for hadrons, depending on the particle direction.

The ATLAS forward calorimeter covers $3.1 < |\eta| < 4.9$ and consists of three modules. The first (electromagnetic) module is a liquid argon/copper calorimeter featuring tubular thin gap electrodes. The two hadronic modules have a tungsten absorber with a similar electrode geometry. The readout of the forward calorimeter is organized in nonprojective rectangular cells, with an approximate cell size of $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$. The total number of calorimeter channels in ATLAS is close to 200,000. And like in CMS, all ATLAS calorimeters are noncompensating as well. For more details, see [72].

5.3.2. Calorimeter jet basics

CMS uses projective calorimeter towers on a grid of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ as input to jet finding. Only towers with $p_T > 1$ GeV/c are considered. The towers correspond to massless 4-vectors by definition, *i.e.*, their kinematic contribution is fully specified by their transverse momentum p_T, their rapidity y, and azimuth ϕ [76]. In ATLAS, two different calorimeter signal definitions are used as input for jet reconstruction. As in CMS, projective cell towers with $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ are used, but without any restriction on the actual value of p_T. This means it is possible to have towers with a nonphysical four-momentum like $E < 0 \Rightarrow p_T < 0$, with η, ϕ, and $m = 0$ fixed by location and/or definition. The negative tower signal is generated by electronic noise as well as by signal fluctuations from the MB (pile-up) events.

On the other hand, the negative energy towers cannot be accepted by the jet finders, as the four-momentum recombination of proto-jets (the towers) requires legitimate, physically allowed kinematic variables. This introduces the need for noise compensation or suppression. In ATLAS, a noise compensation based on the pre-summation of towers into proto-jets has been introduced, which is based on the idea that negative signal towers are merged with neighboring positive signal towers until the total energy in the corresponding proto-jet is above 0. Negative towers without any close-by positive signal towers are dropped.

ATLAS also uses 3-dimensional topological calorimeter cell clusters as input to jet finding. The basic idea of this signal definition is the attempt to reconstruct calibrated *energy blobs* in the calorimeter using the energy flow and shower development correlation between neighboring cells. The result makes optimal use of the fine granularity in ATLAS, especially of the longitudinal segmentation. The clustering algorithm is based on cell signal significance, as measured by the signal-to-noise ratio S/σ, where the noise σ can include pile-up fluctuations. Different thresholds are requested for cells seeding a cluster (primary seeds, $|S/\sigma| > 4$), cells defining the growth of a cluster (secondary seeds, $|S/\sigma| > 2$), and cells to be included because they are direct neighbors of any one of these two seeds ($S/\sigma > 0$). Note that large negative fluctuations can seed a cluster as well. This has been introduced to have an average cancellation of positive noise contributions. Naturally clusters with negative total signal are not used in jet reconstruction, but negative signal cells within a cluster with positive total signal contribute to the jet signal. The algorithm initially puts all topologically connected cells into one cluster. In a

second pass, the clusters are analyzed with respect to local signal maxima. Clusters with more than one local signal maximum are split, and the energy in the cells between the two signal peaks is typically shared by the resulting two clusters [77].

Most clusters have measurable shapes with respect to location, longitudinal and lateral extension, energy sharing in calorimeter compartments, etc. Variables describing these shapes are used to fully calibrate the clusters to a local hadronic energy scale. This procedure includes the attempt to first classify each cluster with respect to the character of its generating particle(s), followed by the application of hadronic weighting functions for "hadronic-looking" clusters. These functions are typically parametrized using shape and location variables (*e.g.*, the depth of cluster center in the calorimeter) sensitive to hadronic-shower development. Next, corrections for dead material losses are applied to all kinds of clusters found near cracks and dead material in the calorimeter. The final cluster correction then attempts to recover signal losses outside of clusters, as introduced by the noise cuts used in the cluster formation algorithm discussed above. Note that none of these procedures use a jet context at all. All calibrations and corrections here are derived from single pion and electron signals alone, and can thus be benchmarked with experimental test-beam data.

Fig. 45 shows a schematic overview on calorimeter jet reconstruction from towers and clusters in ATLAS, respectively. Both calorimeter signal definitions have their specific advantages and disadvantages when used in jet reconstruction. Any particular choice of the calorimeter signal has serious effects on jet reconstruction and calibration, though, as discussed more in the following sections.

5.3.3. Calorimeter jet calibration

Several jet calibration models are under investigation in ATLAS. The most commonly used model is based on a modified cell signal weighting technique, following the original suggestions from H1 [69]. It can be applied to jets from towers as well jets from uncalibrated topological clusters.

In a first step, the cell content of the tower or cluster jet is retrieved. Then, a cell signal weighting function w_c is applied for each cell, depending on its location[11] $\vec{x}_{\text{cell}}$ and the cell signal density $\rho_{\text{cell}} = E_{\text{cell}}/V_{\text{cell}}$, where E_{cell} is the cell signal on an initial (electromagnetic) energy scale, and V_{cell} is the physical volume of the cell. Finally, the jet kinematics is re-calculated using the now calibrated cell signals:

$$P_{\text{jet}} = \sum_{\text{cells}\in\text{jet}} w_c(\vec{x}_{\text{cell}}, \rho_{\text{cell}})\, P_{\text{cell}},$$

with $P_{\text{cell}} = (E_{\text{cell}}, \vec{p}_{\text{cell}})$ and $|E_{\text{cell}}| = |\vec{p}_{\text{cell}}|$. Note that in this scenario the direction of the original jet can change.

The weights can be determined using the particle "truth" jet in simulations by adjusting them such that the reconstructed jet energy on average is identical to the matched particle jet energy. In this scenario the weights reflect all corrections needed to reconstruct the particle jet. In particular, energy lost in particles which do not reach the calorimeter due to the magnetic field is compensated. In the same sense, upstream energy losses in dead material are corrected by this normalization choice.

[11] The cell location is typically indicated by calorimeter module and sampling identifiers, together with a cell index, rather than absolute coordinates, for example.

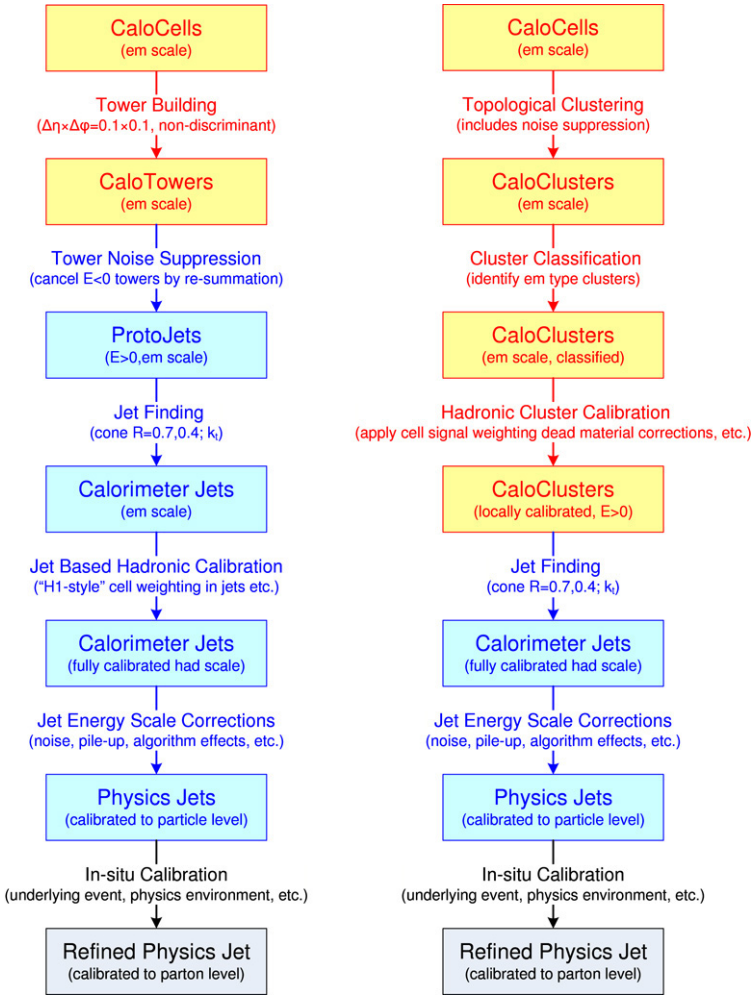


Fig. 45. Jet reconstruction flow in ATLAS. The left diagram shows the sequence for the tower-based reconstruction, while the right diagram shows the cluster-based reconstruction sequence.

Using the calibrated topological clusters as input for jet finding makes the cell signal-based calibration in the jet context, as discussed above, obsolete. In this approach the cluster signals are already fully calibrated in a less biased context, as for example the jet algorithm choice does not enter into the hadronic calibration at this level (see discussion at the end of Section 5.3.2). To first order, the reconstructed jet kinematic is then given by the sum of the calibrated cluster four-momenta. Note that the cluster calibration does not account for energy losses due to the magnetic field and some dead material, as obviously only energy losses with some correlation to the shape or magnitude of a nearby cluster signal can be corrected. Especially energy lost from any cluster, and energy losses due to the loss of small signals by the implicit noise suppression applied in the cell clustering algorithm, still need to be corrected for in the (larger) jet context. An attempt is underway in ATLAS to correct for these effects jet by jet, possibly using measurable jet shapes and detailed cluster information, but not all sensitivities and useful variables are yet fully understood to achieve a particle-level calibration this way.

Important validation signals for jet calibration at the LHC, as at the Tevatron, are prompt photons and hadronic W decays. For systems which have a photon balancing one or more jet(s), corrections can be extracted from the p_T balance. Uncertainties in this procedure mostly arise from initial and final-state radiation, and the underlying event activity, all of which may limit the applicability of these corrections in different collision topologies. The W mass in $W \to q\bar{q}$ can be used for the same purpose. Again, care is required when applying the corrections derived from the mass constraint to other physics topologies. At the LHC, most of the hadronically decaying Ws in the recorded data will be reconstructed in the context of $t\bar{t}$ events, which are very busy final states with potentially large amounts of energy "accidentally" scattered into the jet(s). Also, the W is color-disconnected from the rest of the event, which changes the general p_T flow around its decay jets.

5.3.4. Use of tracks in jets

Reconstructed tracks from the inner tracking detectors in both ATLAS and CMS, in principle, can be used to better calibrate and characterize a given reconstructed jet. Classical energy flow-based reconstruction techniques combine a track with the calorimeter response (typically one cluster) and make use of the feature that the track provides much better energy resolution than the calorimeter for lower energy particles. The application of these techniques promises considerable improvement in the kinematic reconstruction of a single isolated particle like an electron or τ; the application to jets is still under study. In particular, the "subtraction" of the charged response from the calorimeter jet signal is much more challenging due to the already mentioned overlap of showers and generally high tracker occupancy in the jet case. There are indications, however, that the fraction f_{trk} of the jet energy carried by (reconstructed) charged tracks into a jet is already a useful variable to refine the jet calibration, even without individual track/cluster matching. If f_{trk} is large for a given jet, the corresponding calorimeter response has a larger contribution from hadronic showers and may deserve additional calibration corrections to improve the jet energy scale, especially with a precision requirement at the level of 1%.

5.3.5. Jet algorithms

Both ATLAS and CMS use the iterative seeded cone jet finder with $R_{\text{cone}} = 0.4$ (ATLAS) and $R_{\text{cone}} = 0.7$ (ATLAS and CMS), and the k_T jet finder with $D = 0.4$ and $D = 0.6$ (ATLAS), and $D = 1$ in CMS. Other jet algorithms are generally available and under study in particular with respect to certain physics needs (see the discussion in Section 6). The software is implemented such that exactly the same code runs on all possible input objects (partons and hadrons from Monte Carlo generators, detector signals like clusters, towers, reconstructed tracks, etc.), as long as those represent a full four-momentum measure.

5.4. Jet signal characteristics at the LHC

The jet topology as it unfolds in the calorimeters is basically driven by the combination of calorimeter absorption characteristics, the chosen signal definition and jet algorithm choice. This can be seen in the Pythia event in Fig. 46. The tower picture of this particular final state shows rather large jets in general, especially when compared to the hadron jet. This is partly a consequence of the re-summation discussed above, but also due to the fact that towers especially in the endcap and forward regions are filled with signals more generated by the lateral shower development (see jet 4 in Fig. 46) than the hadron energy flow. Topological clusters, on the other hand, collect spatially distributed cell signals, as can again be seen very well in jet 4. In the

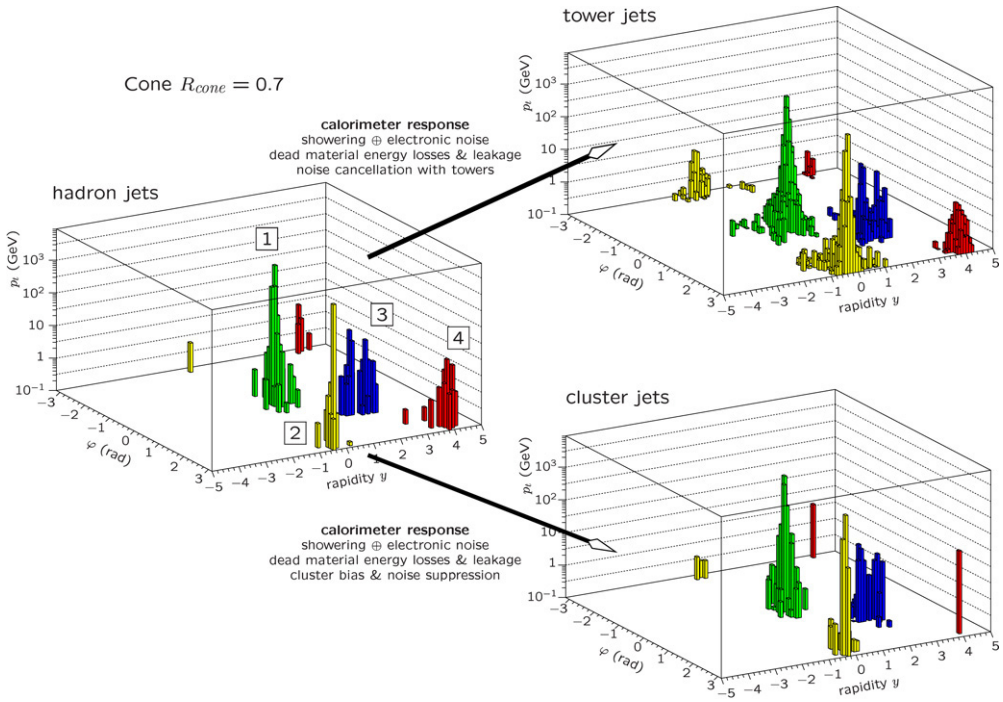


Fig. 46. A simulated QCD 2 → 2 event with two hard cone jets (1 and 2) with energies around 2 TeV each and two softer jets with energies around 20 GeV (3 and 4) in ATLAS. Same colored bins belong to the same jet. The change of the jet shape clearly depends on the calorimeter signal definition.

cluster picture this jet is very collimated in rapidity, and consists of only two clusters. This reflects the rather coarse cell readout granularity of the ATLAS forward calorimeters, which suppresses cluster splitting due to lack of resolvable (lateral) signal structures. In the highly granular central and endcap regions the cluster jet shapes match very well the hadron jet shape, as can be seen from jets 1–3 in this figure.

The particular choice of the calorimeter signal definition to be used in jet reconstruction affects the ability to reconstruct some of the jet kinematics, such as the jet mass. For example, for cone jets with $R = 0.7$ made from calorimeter towers with $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, the number of constituents N_c is given by

$$N_c \approx \frac{\pi R^2_{\text{cone}}}{\Delta\eta \times \Delta\phi = 0.1 \times 0.1} \approx 154,$$

independent of the jet direction and energy. Calorimeter signals such as the topological clusters, which are more sensitive to the hadron-level composition of the jet, typically generate jets with N_c at least indirectly related to the number of incoming particles (see Fig. 47). Naturally, the incoming particle energy flow is convoluted with possibly overlapping shower developments and distributed onto a finite readout granularity. The relation between cell sizes and electromagnetic and hadronic shower sizes puts limitations on the reconstruction of the original incoming particles, and defines the image of the jet in the calorimeter. Fig. 47 indicates that in the central region of ATLAS the cell sizes, even though small in y/η and ϕ, are comparably big with respect

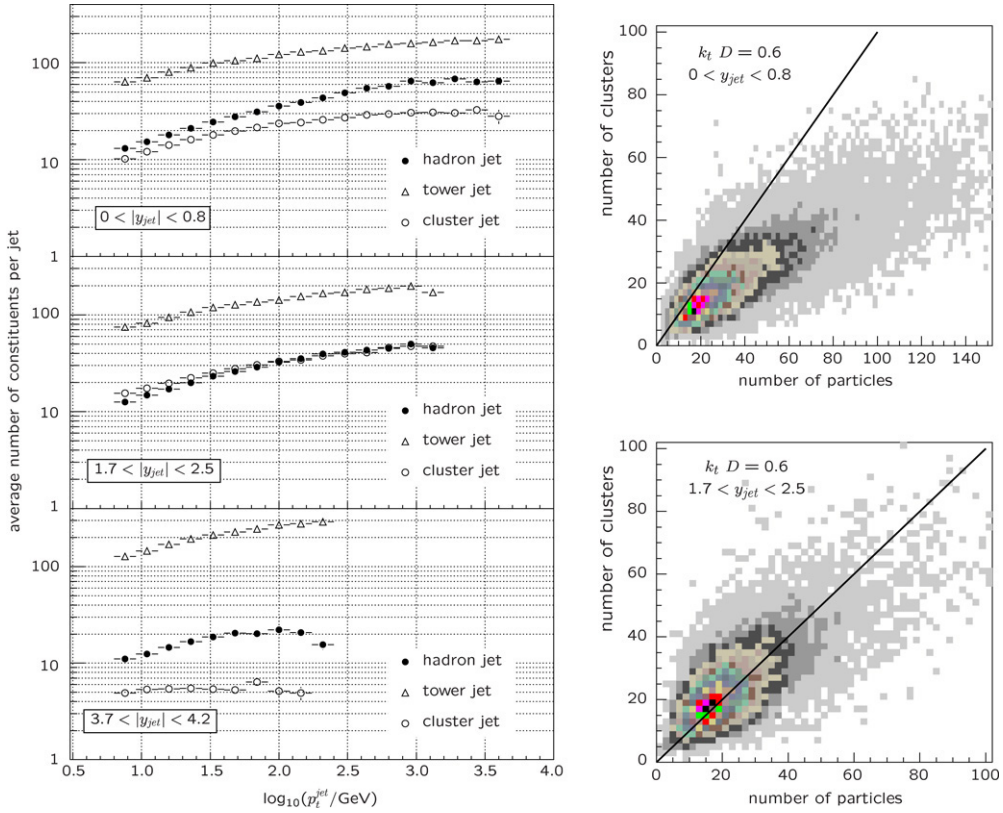


Fig. 47. The number of constituents for hadron, cluster, and tower k_T jets ($D = 0.6$) as a function of the jet p_T for simulated central events in ATLAS, in various regions of rapidity y (left). The right figure shows the number of clusters versus the number of particles in matched cluster and hadron jets in the central and endcap region, from the same simulated data.

to shower sizes, while in the endcap region the cell sizes get sufficiently small, thus improving the structural resolution power of the cluster jets.

One of the more interesting jet variables is its mass. The ability to reconstruct this mass within reasonable precision has considerable impact on the general reconstruction of heavily boosted systems like the top quark, where all final-state decay products may end up within a typical jet cone. In this case the jet mass is the only reconstructable observable giving any indication of the nature of the decaying system. The mass reconstruction is still nearly perfect at the hadronization level, at least if all final-state particles belonging to a given parton are efficiently collected by the chosen jet algorithm. Reconstructing the mass from calorimeter signals is much more challenging in that not only the showering and the resulting signal overlap in the calorimeters smear the mass measurement considerably, but also the fact that the solenoidal fields in front of the calorimeters ($\sim$2 T in ATLAS, $\sim$4 T in CMS) bend charged particles with $p_T < (400\text{–}800)$ MeV$/c$ away from the detectors, *i.e.* outside the jet cone. In addition, the unavoidable amount of dead material typically introduced by the inner detectors and their services as well as calorimeter support structures and cryostat walls (ATLAS only), can significantly reduce the low energy photon signal. This effect has been addressed in a brief study for ATLAS, where the jet mass variation $\delta m/m$ introduced by excluding jet constituents below certain thresholds, starting from 100 MeV

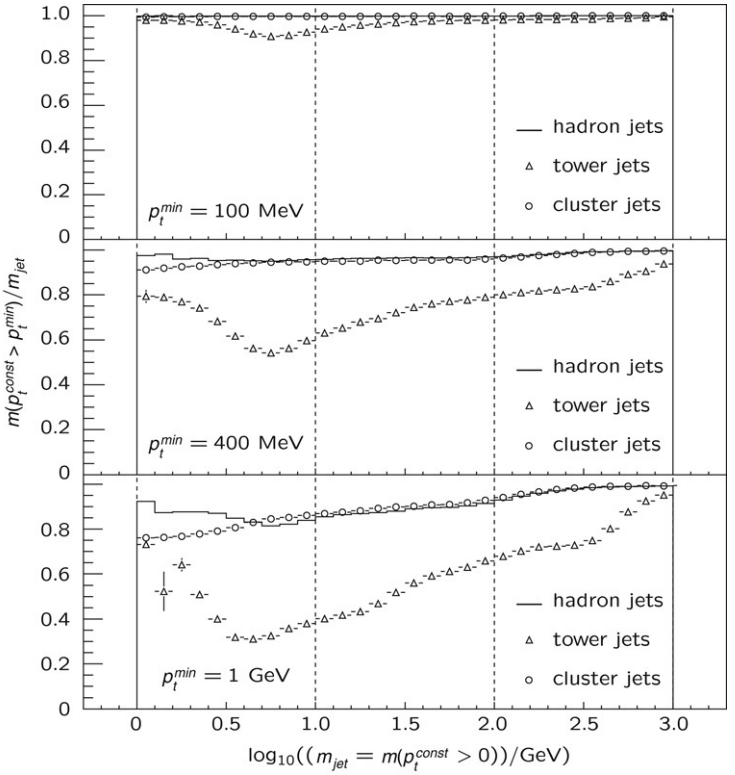


Fig. 48. The change of the jet mass for a constituent selection based on their transverse momentum p_T^{const} ($p_T^{\text{const}} > p_T^{\text{min}}$), for k_T jets from particles, towers, and clusters ($D = 0.6$), as a function of the jet mass calculated from all constituents.

up to 2 GeV, is calculated from QCD di-jet simulations for particles, and the corresponding simulated calorimeter tower and cluster signals with the same cuts applied. Fig. 48 summarizes the results of this study, which indicate that the cluster signals follow the effect at particle level quite well, if compared to tower jets.

Jet reconstruction is therefore in general affected by magnetic field effects, the upstream energy losses from dead materials, showering, leakage, calorimeter regions of low efficiency (cracks), and the underlying and pile-up event activity. Any particular sensitivities of a given jet algorithm to any of these effects can be enhanced or suppressed by a specific calorimeter signal choice, *e.g.* towers or clusters. The jet energy scale calibrations and corrections applied to recover the corresponding energy losses are as much as possible factorized for better control of systematic uncertainties. The magnitude of the contribution of any of these effects to the jet energy scale and jet shape reconstruction varies, depending on the combination of the calorimeter signal choice and the chosen jet algorithm.

One example for an experimentally accessible observable reflecting some of these sensitivities is the jet energy density measure $\psi(r)$, which is the fraction of energy contained in a cone of radius r within a jet. Fig. 49 shows that towers and clusters in the ATLAS calorimeter are expected to produce very similar densities for high p_T jets, but show some differences for jets with $p_T \lesssim 100$ GeV/c. Hadron jets at lower p_T are significantly broader, *i.e.*, have a larger

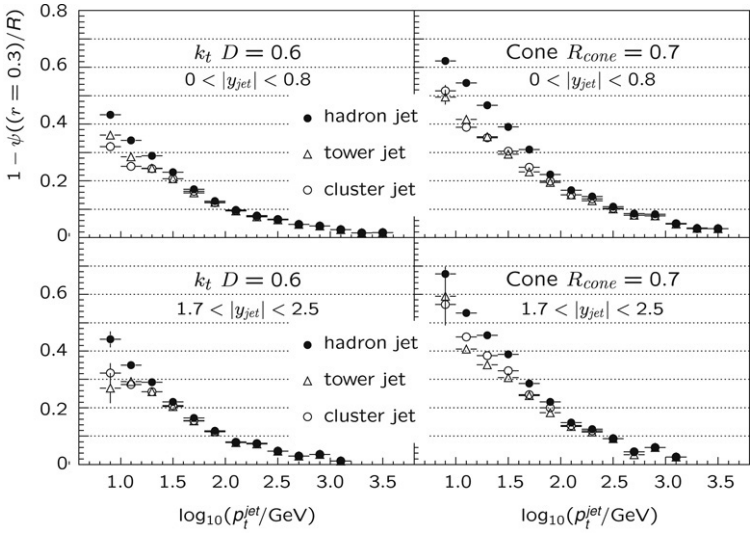


Fig. 49. Estimates from simulations for the ATLAS detector for the energy fraction outside of a cone with $r = 0.3$ around the jet axis, as a function of the jet p_T for k_T (left) and cone (right) jets, in two different regions of rapidity y, for jets in QCD $2 \to 2$ processes.

fraction of the jet energy away from the center of the cone, than either tower or cluster jets. In general the shape of $\psi(p_T)$ in different rapidity regions changes, depending on the jet algorithm choice, in a similar way for all three kinds of jets.

6. SpartyJet

As we have emphasized throughout this review, jets, unlike photons or electrons, are complex objects and the resultant reconstructed 4-vectors may depend on the details of the jet clustering algorithm. Each algorithm has its own strengths and weaknesses and a more robust understanding of the physics of an event can be obtained by examining the result of reconstruction with more than one jet algorithm [78].

Past experience at the Tevatron has been that only one algorithm is typically used for any physics process, at least partially because of the limitations of the analysis machinery. In order to foster a more flexible experimental philosophy, a collection of jet routines was created (SpartyJet) [79] that can allow the reconstruction of the jets from either data or Monte Carlo using multiple algorithms/varied parameters. The routine makes use of the FastKt package for the k_T jet algorithm and the seedless cone algorithm SISCone in addition to other algorithms used by experiments at the Tevatron and LHC. The program can run either in the ROOT [80] format, for example inside an ntuple, or it can run in stand-alone fashion on a collection of 4-vectors. As an example, in Fig. 50, is shown the results of running SpartyJet on a sample of Monte Carlo events generated for the ATLAS experiment. The sample consists of di-jet events with a $p_{T,\mathrm{min}}$ of approximately 2 TeV/c and the inputs to SpartyJet are the topological calorimeter clusters discussed in the previous section. The cone algorithms are run with a cone radius of 0.7 and a split/merge criterion of 0.75. The k_T algorithms are run with a D parameter of 0.7 as well. The di-jet character of the events can be seen by the clear peak at approximately 2 TeV/c, and the impact of hard gluon radiation off the initial and final states can be observed in the sizable

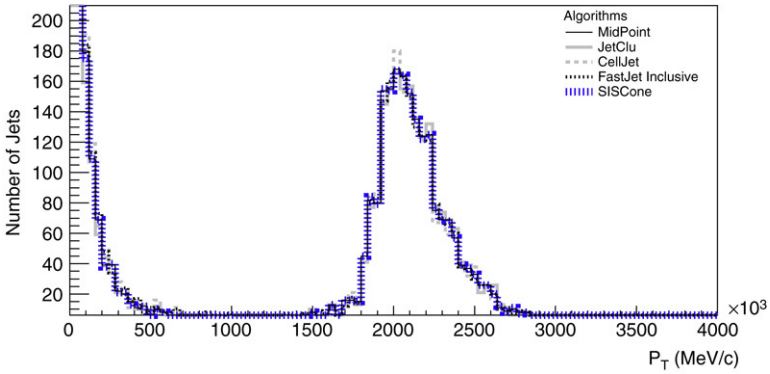


Fig. 50. The inclusive jet cross-section for the LHC with a $p_{T,\min}$ value for the hard scattering of approximately 2 TeV/c, using several different jet algorithms with a distance scale ($D = R_{\text{cone}}$) of 0.7. The first bin has been suppressed.

lower tail at lower transverse momenta. The jets found by the second pass algorithm for the midpoint algorithm can also be observed at the lower transverse momentum values. Note that on this plot the cone and k_T algorithms with similar scale parameters give similar results for the cross-section. Any differences need more detailed comparisons to become apparent.

Another interesting variable to plot using SpartyJet is the distribution of jet masses. As discussed in Section 3.3, the mass of typical QCD (gluon or light quark) jets is generated primarily by perturbative gluon emission. This is to be contrasted with a jet arising from a heavy quark (such as a top quark) that also has an intrinsic mass from the heavy quark. The distribution of jet masses from a typical QCD event sample (restricting the transverse momentum range to 1.8–2.2 TeV/c) is shown in Fig. 51. There is a Sudakov suppression of low jet masses, which can arise only if there is little or no gluon radiation from the short-distance final-state partons. At jet masses above the peak (here at approximately 125 GeV/c^2) the jet mass distribution falls slowly, roughly between $1/m$ and $1/m^2$, with the average jet mass at a value above the peak (approximately 150 GeV/c^2 in this sample). There is also a suppression of jets with large masses due to the tendency of the jet algorithms to split jets in which the energy is widely dispersed. Note that for these very high p_T jet events that a jet mass of the order of 175 GeV/c^2 (purely from gluon emission) is not uncommon, so this is a caveat for the naive use of the mass of a jet to search for highly boosted top quarks.

One can gain further information about the origin of a jet by examining the y-scale[12] (using the k_T algorithm) at which the jet can be split into two subjets. This ability has been implemented into SpartyJet by the use of the y-splitter routine [81]. This scale will tend to be larger for highly boosted massive objects (like a high p_T W or top quark) than for QCD jets at the same transverse momentum. For example, for a boosted W, the y scale for resolving the W-jet into two subjets should be on the order of $m_W/p_{T,\text{jet}}$, while for jet structures produced by QCD radiation, the scales should be much smaller. The y-scale distribution for the jets from the 1.8–2.2 TeV/c jet sample to be split into two subjets is shown in Fig. 52. Low scales dominate as expected for a QCD jet sample.

The average jet mass is plotted in Fig. 53 versus the transverse momentum of the jet for several jet algorithms for an inclusive jet Monte Carlo sample with transverse momenta from

[12] $y_2 = \min(p_{T,1}^2, p_{T,2}^2) \cdot d_2^2/p_{T,\text{jet}}^2$, where $p_{T,1}^2$ and $p_{T,2}^2$ are the transverse momentum values of the two subjets and d_2 is the distance scale at which the jet is divided into two subjets.

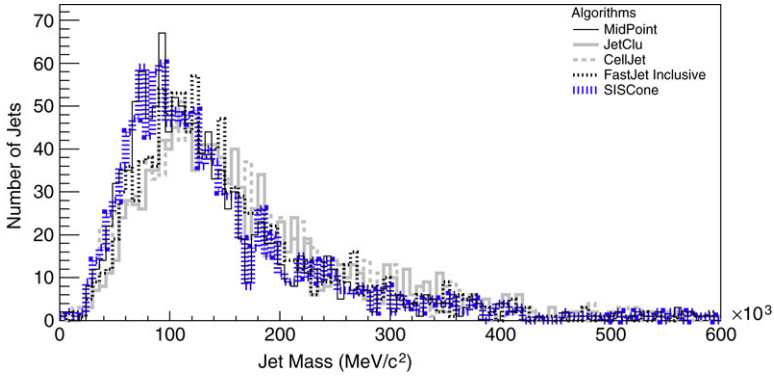


Fig. 51. The jet mass distributions for an inclusive jet sample generated for the LHC with a $p_{T,\min}$ value for the hard scattering of approximately 2 TeV/c, using several different jet algorithms with a distance scale ($D = R_{\text{cone}}$) of 0.7.

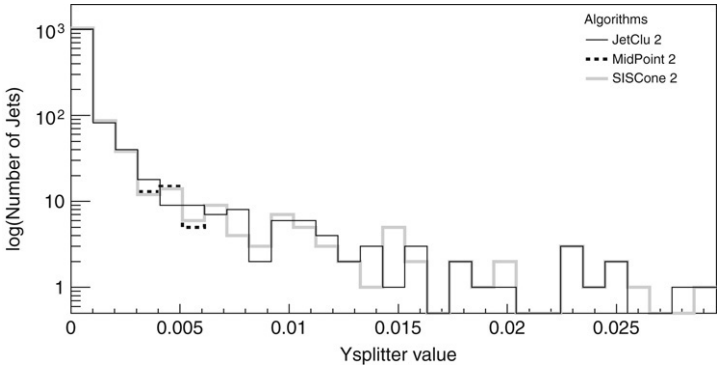


Fig. 52. The y-scale distributions for the jets from the 1.8–2.2 TeV/c jet sample to be split into two subjets. Several different jet algorithms with a distance scale ($D = R_{\text{cone}}$) of 0.7 are used.

100 GeV/c to 1 TeV/c. The average mass increases roughly linearly with the jet transverse momentum. The reconstructed Monte Carlo jet mass values are in reasonable agreement with the NLO perturbative predictions discussed in Section 3.3. Thus, the jet mass, just as the jet shape, can be reasonably described by a NLO partonic calculation. The average Monte Carlo jet mass at high p_T tends to approach the NLO prediction carried out with the use of an R_{sep} value of 1 (compared to the canonical value of 1.3), perhaps due to the impact of the very narrow jet profiles at these high transverse momenta. All of the jet algorithms result in a similar average jet mass, although JetClu tends to give larger results, due to the effects of ratcheting. As discussed in Section 3.3, the average jet mass is expected to scale as $R \cdot p_T$, where R is the size parameter for the jet. It is interesting to note that variations in theoretical (R_{sep}) and in experimental jet algorithms (ratcheting) have little impact on the magnitude of the inclusive jet cross-section, but do have a noticeable effect on the jet mass.

A lego plot of a single event from the high p_T jet sample is shown in Fig. 54(a). Again, there is a clear di-jet structure in which the jets have very collimated cores but the presence of such a large scale in the event results in there being a several extra jets of quite sizable transverse momentum in their own right. This becomes more apparent when we change the transverse momentum scale

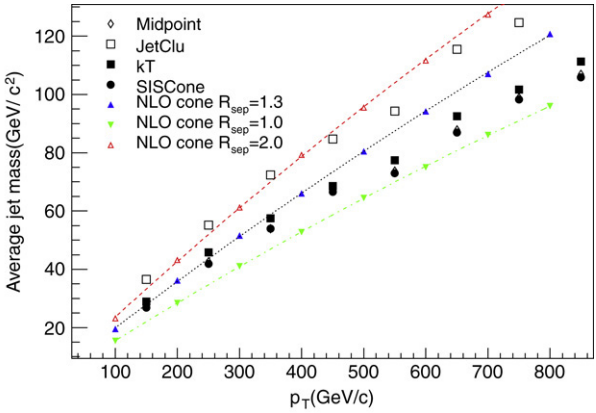


Fig. 53. The average jet mass is plotted versus the transverse momentum of the jet using several different jet algorithms with a distance scale ($D = R_{\text{cone}}$) of 0.7.

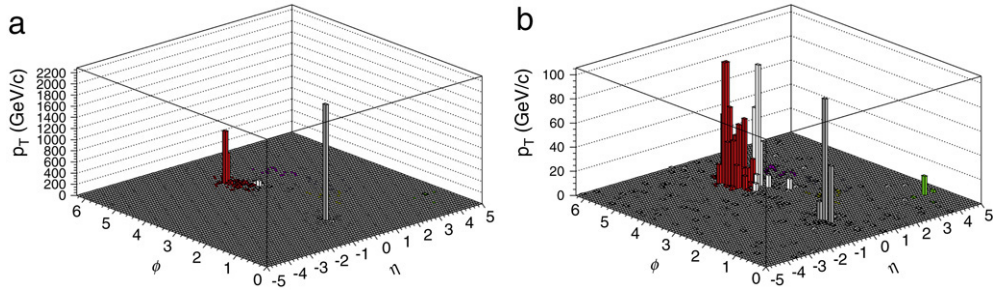


Fig. 54. (a) A lego plot for an event from the inclusive jet sample generated for the LHC with a $p_{T,\min}$ value for the hard scattering of approximately 2 TeV/c. (b) A lego plot for the same event; however, the p_T scale has been cut off at 100 GeV/c to give a clear view of low p_T jets.

in the lego plot as shown Fig. 54(b). Note that the jet colored in white is a second pass jet, having been missed by the midpoint algorithm in the first pass due to the effects discussed earlier.

SpartyJet is currently in use by ATLAS, CDF and CMS.

7. Conclusions

Jets are present in nearly every final state measured at the Tevatron. This will be true as well at the LHC, with the environment for jet measurements being even more challenging. Thus, it is crucial both to improve our understanding of the subtleties of jet reconstruction as well as to continue the development of new tools. In this article, we have tried to remove some of the mystique regarding the measurement of jets at hadron–hadron colliders by pointing out the connections between the theoretical predictions and the experimental measurements, and the similarities and differences between the cone and k_T algorithms. In the process we have reviewed the history of jets at the Tevatron, including both the experimental and theoretical successes, and the surprises and mis-steps. Looking to the future we have outlined the issues expected to be important at the LHC. Specifically, we have discussed jet reconstruction and jet energy calibration experience at the Tevatron experiments and the on-going work at LHC experiments.

LHC experiments benefit from the experience at the Tevatron (and also other experiments such as the HERA experiments), and, due to excellent capabilities of the ATLAS and CMS calorimeters, more advanced schemes are being explored such as the use of topological calorimeter cell clusters as inputs to jet clustering. An essential payoff of measurements performed at the Tevatron is the ability to accurately tune Monte Carlo event simulations to improve the modeling of jets. As we have discussed it is important that these experimental measurements be presented at the hadron level to facilitate the model tuning process. These measurements must also be repeated in the new environment of the LHC experiments.

In this context of past experiences and future expectations we have made several recommendations that we feel will play an essential role in the successful analysis of the data from the LHC. These include:

- the use of a variety of jet algorithms for physics analyses with continuous cross-checking of results
- the use of 4-vector kinematics, including evaluation of the jet mass, to characterize a jet
- the use of seedless algorithms (or correction back to seedless) in cone-based jet clustering
- the correction (where possible) of jets back to the hadron level in experimental analyses.

In addition, we have presented a framework (SpartyJet) that facilitates the use of multiple jet algorithms in both experimental and theoretical studies. We close by applauding the 20 years of highly successful jet physics at the Tevatron and looking forward to an equally exciting application of jets to physics beyond the Standard Model at the LHC.

Acknowledgements

We would like to thank Anwar Bhatti, Albert de Roeck, Lance Dixon, Rick Field, Kurtis Geerlings, Craig Group, Aurelio Juste, Christophe Royon, Gavin Salam and Markus Wobisch for useful conversations and supply of figures. We also would like to thank the members of the ATLAS jet performance working group and the members of the ATLAS simulation production team for providing the simulated data for this experiment. In particular, we are grateful for help from Walter Lampl, Chiara Paleari and Rolf Seuster, especially concerning ATLAS jet reconstruction software issues.

References

[1] J.M. Campbell, J.W. Huston, W.J. Stirling, Rep. Prog. Phys. 70 (2007) 89. arXiv:hep-ph/0611148.
[2] A.A. Affolder, et al., (CDF Collaboration), Phys. Rev. D 65 (2002) 092002.
[3] See, e.g. J. Pumplin, et al., JHEP07 012 (2002). arXiv:hep-ph/0201195;
R.S. Thorne, A.D. Martin, W.J. Stirling. arXiv:hep-ph/0606244.
[4] See, e.g. T. Sjostrand, S. Mrenna, P. Skands, JHEP05 026 (2006) (LU TP 06-13, FERMILAB-PUB-06-052-CD-T). arXiv:hep-ph/0603175.
[5] See, e.g. G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, B.R. Webber, JHEP 0101 (2001) 010. arXiv:hep-ph/0011363; hep-ph/0210213.
[6] See also the summary in the recent Tevatron-for-LHC Report of the QCD Working Group, FERMILAB-CONF-06-359, Oct. 2006. arXiv:hep-ph/0610012.
[7] See, e.g. J.C. Collins, A. Metz, Phys. Rev. Lett. 93 (2004) 252001. arXiv:hep-ph/0408249, and references therein.
[8] See the discussion in Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop, G.C. Blazey, et al. arXiv:hep-ex/0005012.
[9] J.E. Huth, et al., in: E.L. Berger (Ed.), Proceedings of Research Directions for the Decade: Snowmass 1990, World Scientific, Singapore, 1992, p. 134.

[10] S.D. Ellis, Z. Kunszt, D. Soper, Phys. Rev Lett. 69 (1992) 1496; Phys. Rev. Lett. 64 (1990) 2121; Phys. Rev. D 40 (1989) 2188; Phys. Rev. Lett. 62 (1989) 726.
[11] F. Abe, et al., (CDF Collaboration), Phys. Rev. D 45 (1992) 1448;
B. Abbott, et al., (DØ Collaboration), FERMILAB-Pub-97/242-E.
[12] F. Abe, et al., (CDF Collaboration), Phys. Rev. Lett. 68 (1992) 1104.
[13] F. Abe, et al., (CDF Collaboration), Phys. Rev. Lett. 70 (1993) 713.
[14] S.D. Ellis, Z. Kunszt, D. Soper, Phys. Rev. Lett. 69 (1992) 3615. arXiv:hep-ph/9208249;
S.D. Ellis, Proceedings of the 28th Rencontres de Moriond: QCD and High Energy Hadronic Interactions, March 1993, p. 235;
B. Abbott, et al., Fermilab Pub-97-242-E, 1997.
[15] M.H. Seymour, Nuclear Phys. B 513 (1998) 269. arXiv:hep-ph/9707338.
[16] Gavin P. Salam, Gregory Soyez. arXiv:hep-ph/0704.0292.
[17] G. Salam, private communication.
[18] S.D. Ellis, D. Soper, Phys. Rev. D 48 (1993) 3160. arXiv:hep-ph/9305266.
[19] S. Catani, Yu.L. Dokshitzer, B.R. Webber, Phys. Lett. B 285 (1992) 291;
S. Catani, Yu.L. Dokshitzer, M.H. Seymour, B.R. Webber, Nuclear Phys. B 406 (1993) 187.
[20] M. Cacciari, G.P. Salam, Phys. Lett. B 641 (2006) 57. arXiv:hep-ph/0512210.
[21] Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, JHEP 9708 (1997) 001. arXiv:hep-ph/9707323.
[22] M. Wobisch, T. Wengler. arXiv:hep-ph/9907280.
[23] J.M. Butterworth, J.P. Couchman, B.E. Cox, B.M. Waugh, Comput. Phys. Commun. 153 (2003) 85. arXiv:hep-ph/0210022; see also http://hepforge.cedar.ac.uk/ktjet/.
[24] S.D. Ellis, J. Huston, M. Tönnesmann. arXiv:hep-ph/0111434.
[25] D. Acosta, et al., (CDF Collaboration), Phys. Rev. D 71 (2005) 112002. arXiv:hep-ex/0505013.
[26] See, e.g., the overview given by R. Field, The CDF and DØ Collaborations, FERMILAB-CONF-05-596-E. Published Proceedings 35th International Symposium on Multiparticle Dynamics, ISMD 05, Kromeriz, Czech Republic, 9–15 August, 2005.
[27] F. Abe, et al., [CDF Collaboration], Phys. Rev. Lett. 73 (1994) 225. arXiv:hep-ex/9405005.
[28] F. Abe, et al., [CDF Collaboration], Phys. Rev. Lett. 74 (1995) 2626. arXiv:hep-ex/9503002.
[29] S. Abachi, et al., [D0 Collaboration], Phys. Rev. Lett. 74 (1995) 2632. arXiv:hep-ex/9503003.
[30] M. Carena et al. [Higgs Working Group Collaboration]. arXiv:hep-ph/0010338.
[31] CDF and DØ Working Group members, L. Badukhadia et al., FERMILAB-PUB-03-320-E, 2003.
[32] A. Juste, C. Royon, private communication.
[33] A. Bhatti, et al., Nucl. Instrum. Methods A 566 (2006) 375. arXiv:hep-ex/0510047.
[34] W.T. Giele, E.W.N. Glover, D.A. Kosower, Phys. Rev. Lett. 73 (1994) 2019. arXiv:hep-ph/9403347.
[35] Z. Nagy, Phys. Rev. Lett. 88 (2002) 122003. arXiv:hep-ph/0110315; Phys. Rev. D 68 (2003) 094002. arXiv:hep-ph/0307268.
[36] S.D. Ellis. arXiv:hep-ph/9306280.
[37] S. Frixione, B.R. Webber, JHEP 0206 (2002) 029. arXiv:hep-ph/0204244.
[38] B. Abbott, et al., [D0 Collaboration], Nucl. Instrum. Methods A 424 (1999) 352. arXiv:hep-ex/9805009.
[39] G. Grindhammer, M. Rudowicz, S. Peters, Nucl. Instrum. Methods A 290 (1990) 469.
[40] F. Abe, et al., [CDF Collaboration], Phys. Rev. Lett. 80 (1998) 5720. arXiv:hep-ex/9711004.
[41] A. Abulencia, et al., [CDF Collaboration], Phys. Rev. Lett. 96 (2006) 022004. arXiv:hep-ex/0510049.
[42] V.M. Abazov, et al., [DØCollaboration], Phys. Rev. D 74 (2006) 092005. arXiv:hep-ex/0609053.
[43] A. Abulencia, et al., [CDF Collaboration], Phys. Rev. D 74 (2006) 071103. arXiv:hep-ex/0512020.
[44] A. Abulencia, et al., [CDF Collaboration], Phys. Rev. Lett. 96 (2006) 122001. arXiv:hep-ex/0512062.
[45] A. Abulencia et al.CDF Collaboration. arXiv:hep-ex/0701051. [56].
[46] C. Group, Ph.D. Thesis, University of Florida, 2006.
[47] V.M. Abazov, et al., [D0 Collaboration], Phys. Rev. Lett. 94 (2005) 221801. arXiv:hep-ex/0409040.
[48] T. Gleisberg, S. Hoche, F. Krauss, A. Schalicke, S. Schumann, J.C. Winter, JHEP 0402 (2004) 056. arXiv:hep-ph/0311263.
[49] M.L. Mangano, M. Moretti, R. Pittau, Nuclear Phys. B 632 (2002) 343. arXiv:hep-ph/0108069.
[50] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, JHEP 0307 (2003) 001. arXiv:hep-ph/0206293.
[51] D. Acosta, et al., [CDF Collaboration], Phys. Rev. D 70 (2004) 072002. arXiv:hep-ex/0404004.
[52] R.D. Field for the CDF Collaboration, FERMILAB-CONF-06-409-E, Presented at 33rd International Conference on High Energy Physics, ICHEP 06, Moscow, Russia, 26 July–2 August 2006.
[53] R. Field, “Herwig, Jimmy and Pythia Tune A”, talk given at TeV4LHC at Fermilab, 2004.

[54] M. Wobisch, private communication.
[55] D. Stump, J. Huston, J. Pumplin, W.K. Tung, H.L. Lai, S. Kuhlmann, J.F. Owens, JHEP 0310 (2003) 046. arXiv:hep-ph/0303013.
[56] H.L. Lai, et al., [CTEQ Collaboration], Eur. Phys. J. C 12 (2000) 375. arXiv:hep-ph/9903282.
[57] M. Voutilainen [D0 Collaboration], Proceedings of the DIS2006 Workshop. arXiv:hep-ex/0609026.
[58] A.D. Martin, R.G. Roberts, W.J. Stirling, R.S. Thorne, Phys. Lett. B 604 (2004) 61. arXiv:hep-ph/0410230.
[59] T. Kluge, K. Rabbertz, M. Wobisch. arXiv:hep-ph/0609285.
[60] J.M. Campbell, R.K. Ellis, Phys. Rev. D 62 (2000) 114012. arXiv:hep-ph/0006304.
[61] S. Mrenna, P. Richardson, JHEP 0405 (2004) 040. arXiv:hep-ph/0312274.
[62] D. Acosta, et al., [CDF Collaboration], Phys. Rev. D 71 (2005) 052003. arXiv:hep-ex/0410041.
[63] C. Neu, [CDF Collaboration], FERMILAB-CONF-06-162-E.
[64] A. Abulencia, et al., [CDF Collaboration], Phys. Rev. D 74 (2006) 072006. arXiv:hep-ex/0607035.
[65] D. Acosta, et al., [CDF Collaboration], Phys. Rev. D 72 (2005) 032002. arXiv:hep-ex/0506001.
[66] U. Aglietti, et al. arXiv:hep-ph/0612172.
[67] CDF Collaboration, Determination of the energy scale of b-jets using the $Z \to b\bar{b}$ signal, CDF public note 8746.
[68] DØ Collaboration, Evidence for $Z \to b\bar{b}$ Decays at DØ, DØnote 5205-CONF.
[69] C. Adloff, et al., [H1 Collaboration], Z. Phys. C 74 (1997) 221. arXiv:hep-ex/9702003.
[70] G. Abbiendi, et al., [OPAL Collaboration], Eur. Phys. J. C 12 (2000) 567. arXiv:hep-ex/9908002.
[71] S. Lami, A. Bocci, S. Kuhlmann, G. Latino, [CDF Collaboration], Frascati Phys. Ser. 21 (2001) 555.
[72] ATLAS Coll., ATLAS detector and physics technical design report, CERN-LHCC-99-14/15, 1999.
[73] A. Moraes, C. Buttar, D. Clements, Measuring the underlying event at ATLAS, ATL-PHYS-PUB-2005-015, 2005.
[74] R.A. Davis, P. Savard, A study of pileup noise in the barrel and endcap calorimetry ATL-CAL-96-084, 1996.
[75] CMS Coll., CMS physics technical design report, vol.1, CERN-LHCC-2006-001, 2006.
[76] CMS Coll., J. Phys. G: Nucl. Part. Phys. 34 (2007).
[77] S. Menke, private communication.
[78] M.H. Seymour, C. Tevlin, JHEP 0611 (2006) 052. arXiv:hep-ph/0609100.
[79] See: www.pa.msu.edu/˜huston/SpartyJet/SpartyJet.html.
[80] See: root.cern.ch.
[81] J.M. Butterworth, B.E. Cox, J.R. Forshaw, Phys. Rev. D 65 (2002) 096014. arXiv:hep-ph/0201098.