

## Snowmass 2001: Jet Energy Flow Project\*

C.F. Berger,<sup>1</sup> E.L. Berger,<sup>2</sup> P.C. Bhat,<sup>3</sup> J.M. Butterworth,<sup>4</sup> S.D. Ellis,<sup>5</sup> B. Flaughner,<sup>3</sup>  
 W.T. Giele,<sup>3</sup> W. Kilgore,<sup>6</sup> A. Kulesza,<sup>6</sup> S. Lammers,<sup>7</sup> S. Magill,<sup>2</sup> and H. Prosper<sup>8</sup>

<sup>1</sup>*State University of New York, Stony Brook*

<sup>2</sup>*Argonne National Laboratory*

<sup>3</sup>*Fermi National Accelerator Laboratory*

<sup>4</sup>*University College London*

<sup>5</sup>*University of Washington*

<sup>6</sup>*Brookhaven National Laboratory*

<sup>7</sup>*University of Wisconsin/DESY*

<sup>8</sup>*Florida State University*

(Dated: February 19, 2002)

### Abstract

Conventional cone jet algorithms arose from heuristic considerations of LO hard scattering coupled to independent showering. These algorithms implicitly assume that the final states of individual events can be mapped onto a unique set of jets that are in turn associated with a unique set of underlying hard scattering partons. Thus each final state hadron is assigned to a unique underlying parton. The Jet Energy Flow (JEF) analysis described here does not make such assumptions. The final states of individual events are instead described in terms of flow distributions of hadronic energy. Quantities of physical interest are constructed from the energy flow distribution summed over all events. The resulting analysis is less sensitive to higher order perturbative corrections and the impact of showering and hadronization than the standard cone algorithms.

---

\* Contribution to the P5 Working Group on QCD and Strong Interactions at Snowmass 2001

## A. Introduction

The goal of a jet algorithm is to provide a precise mapping between the observed, long distance hadronic final states in high energy interactions and the underlying energetic partons participating in the true short-distance, hard scattering process[1]. To appreciate this goal imagine comparing the observed final states detected by “real” detectors that have sizes of order centimeters to meters to what would be observed with a detector whose size is characterized by a distance scale of a fraction of a fermi. Between these two scales, *i.e.*, from a fraction of a fermi to centimeters, the short-distance state evolves via higher order perturbative processes and the physics associated with showering and hadronization. These processes allow the short-distance partons, along with the spectators, to evolve into the observed hadrons. During the evolution the 4-momentum associated with an initial short-distance parton is spread out over a number of final state hadrons occupying an extended region of phase space. To achieve our stated goal, the jet algorithm tries to identify these “related” hadrons into a single jet, whose total 4-momentum should track that of the initial parton. To perform this task with precision it is important that the results of applying the jet algorithm are insensitive to both higher order corrections and fluctuations in the showering/hadronization processes. The results of the jet algorithm should also be insensitive to the smearing effects of the detection process itself, *e.g.*, due to the granularity of the detector.

Current jet algorithms attempt to achieve the stated goal in a quite singular way by assigning the observed hadrons to *unique* jets on an *event-by-event* basis. This identification proceeds in the face of the fundamental fact that such a unique assignment cannot be precise. While the underlying parton that initiates the jet is a nonsinglet under the color symmetry of QCD, the hadrons are all color singlets. At the very least, the jets in the final state must represent the correlated evolution of more than one short-distance parton or of a short-distance parton correlated with a spectator parton, *i.e.*, a color singlet set of initial partons. In some sense the most extreme approach is represented by cone algorithms as advocated at the 1990 Snowmass Workshop[2]. Cone algorithms associate hadrons into jets by identifying those that are nearby in angle. The underlying assumption is that the extra radiation produced by higher order corrections, showering and hadronization around an energetic parton appears symmetrically, *i.e.*, occurs independently of the other color

charged objects in the final state. The alternative jet algorithm, the  $K_T$  algorithm[3], uses nearness in momentum space to identify the members of a jet and, at least to some extent, recognizes the “color-connectedness” of the radiation producing the final state. However, as noted above, both of these algorithms associate individual hadrons with unique jets on an event-by-event basis. We know that this procedure is only an approximation (due to the color conservation issue) and can lead to undesirable dependence (at least at the 10% level) on the details of the showering and hadronization processes. In the context of cone jet algorithms this latter point is discussed in more detail in another contribution to these Proceedings[4].

The Jet Energy Flow (JEF) approach described in this note is a simplified version of the more completely developed C-continuous observables or C-algebra formalism of F. V. Tkachov[5] for describing energy flow in hadronic collisions[6]. JEF accepts the reality that the hadronic final state represents the collective radiation from several out-flowing color charges, *i.e.*, the underlying short-distance partons. No attempt is made to associate individual hadrons with unique jets, *i.e.*, with unique underlying partons, on an event-by-event basis. Yet the energy flow pattern of an event still provides a footprint of the underlying partons, from which much of the same information provided by the standard jet algorithms can be extracted. As the subsequent discussion will indicate, it is a more reliable characterization of the event in the sense of exhibiting a reduced sensitivity to the showering and hadronization processes. The challenge in the JEF type analysis is to define observables that offer an informative comparison between theory and experiment.

## **B. The JEF Formalism**

A primary strength of the JEF approach is that, in contrast with the usual algorithmic approach to jet identification, the JEF formalism generates, event-by-event, a smooth distribution to characterize each event. In that sense, the JEF formalism is more analytic. For example, in the application of the cone algorithm the goal of identifying unique jets leads to the “stability” constraint[4]. A set of hadrons or partons that lie within a cone of a defined size  $R$  are identified as constituting a jet if and only if the energy-weighted centroid of the set of particles coincides with the geometric center of the cone. This constraint results in the non-analytic structure of the implementation of the algorithm, typically in the form

of step functions with complicated arguments. Only limited (and typically complicated) regions of the multi-particle phase space contribute to a jet. No such constraint arises in the JEF analysis. This distinction has several important consequences.

1. The more inclusive and analytic calculations characteristic of a JEF analysis are more amenable to resummation techniques and power corrections analysis in perturbative calculations.
2. Since the required multi-particle phase space integrations are largely unconstrained, *i.e.*, more analytic, they are easier (and faster) to implement. Programs like JETRAD spend considerable computer time simulating the complicated phase space required by the algorithmic style jet algorithms.
3. Since the analysis does not identify jets event-by-event, the analysis of the experimental data from an individual event should proceed more quickly.
4. Signal to background optimization can now include the JEF parameters (and distributions). One cannot typically optimize a standard jet algorithm except for a limited number of parameters.

We can define the fundamental distribution of the JEF analysis as follows. We start with a set of 4-vectors,  $p_\mu = (E, p_x, p_y, p_z)$ , that represent either the partons in a perturbative calculation or the hadrons in a simulated or real event. In the latter case these 4-vectors might correspond as well to the location and energy deposited in individual calorimeter cells. If a given event corresponds to the measurement of  $N$  such 4-vectors,  $\{p_\mu^i\}_{i=1}^N = \left\{ \left( E^i, \vec{P}^i \right) \right\}_{i=1}^N$ , we have the 4-vector distribution for that event defined by

$$P_\mu \left( \hat{P} \right) = \sum_{i=1}^N p_\mu^i \delta \left( \hat{P} - \hat{P}^i \right), \quad (1)$$

where the directional unit vector is defined by  $\hat{P}(m) = \vec{P} / |\vec{P}|$  with the 2-dimensional angular variable defined as  $m = (\theta, \phi)$  (typical of lepton colliders) or  $m = (\eta, \phi)$  (typical of hadron colliders, where  $\eta$  is the pseudorapidity,  $\eta = \ln(\cot \theta/2)$ ). This expression defines the underlying energy flow via  $E(m) = P_0(m)$  with a corresponding expression for the underlying longitudinal momentum flow  $P_z(m)$ . For the case of hadronic colliders

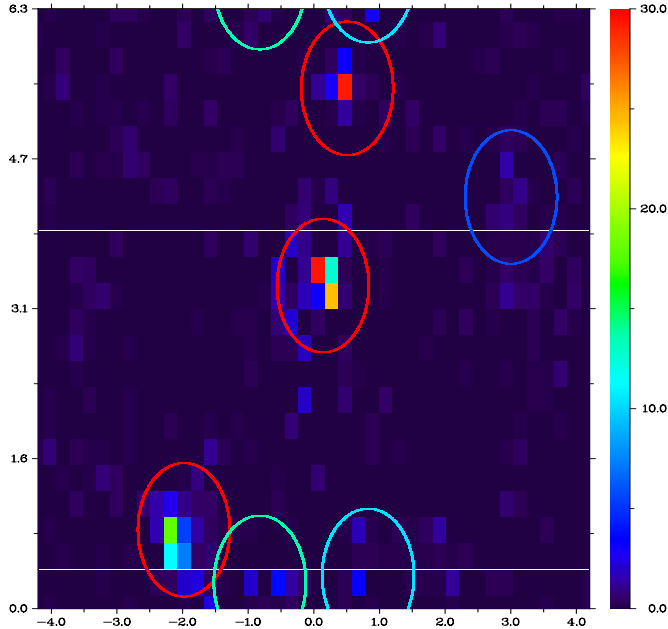


FIG. 1: Transverse energy (in GeV color coding) calorimeter readout in a typical CDF event as a function of pseudorapidity and azimuthal angle. The reconstructed CDF cone jets are indicated by the circles (with the same GeV color coding).

the more familiar underlying transverse energy flow is defined by the composite quantity  $E_T(m) = \sqrt{P_x^2(m) + P_y^2(m)}$  or  $E_T(m) = E(m) \times \sin \theta(m)$ . Clearly many quantities can be constructed from the 4-momentum distribution of Eq. 1, including the usual cone jet algorithm. The  $E_T(\eta, \phi)$  distribution for a typical CDF jet event is illustrated in Figure 1 along with the cone jets “found” with the CDF cone jet algorithm.

Using the underlying 4-vector distribution we define the jet energy flow (JEF) via a smearing or averaging function  $A$  as

$$J_\mu(m) \equiv \int dm' P_\mu(m') \times A(m' - m), \quad (2)$$

where  $A$  is normalized as

$$\int dm A(m) = 1. \quad (3)$$

A simple (but not unique) form for the averaging function in terms of the general 2-tuple of angular variables  $m = (\alpha, \beta)$ , which provides a direct comparison with the jet cone

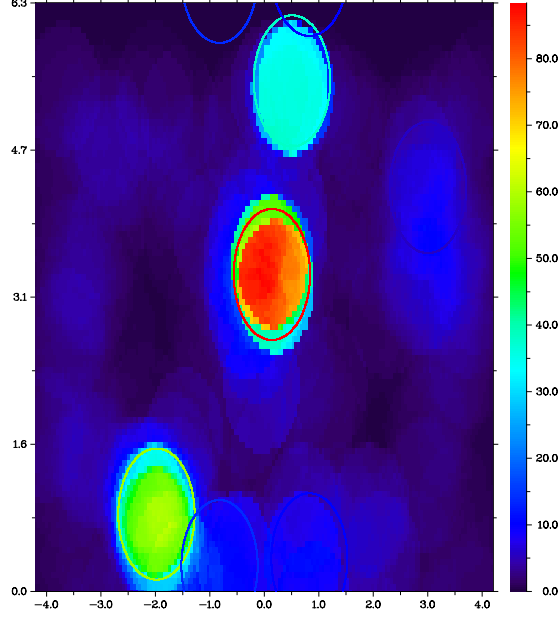


FIG. 2: The  $E_T(\eta, \phi)$  flow (including the factor  $\pi R^2$ ) using the CDF event of Figure 1.

algorithm, is

$$A(m) = A(\alpha, \beta) = \frac{\Theta(R - r(\alpha, \beta))}{\pi R^2} = \frac{\Theta\left(R - \sqrt{\alpha^2 + \beta^2}\right)}{\pi R^2}, \quad (4)$$

where  $R$  is the cone size and  $r(\alpha, \beta)$  is the distance measure in the space defined by  $(\alpha, \beta)$ . For comparison with the existing jet cone analyses we will discuss the case  $m = (\eta, \phi)$ . As a specific example we exhibit the jet transverse energy flow (transverse JEF)

$$J_T(m) = \int dm' E_T(m') \times A(m' - m) = \int dm' \sqrt{P_x^2(m') + P_y^2(m')} \times A(m' - m) \quad (5)$$

times a factor of  $\pi R^2$  ( $E_T = \pi R^2 \times J_T$ ) for the case of  $R = 0.7$  in Figure 2 and 3 for the same event displayed in Figure 1. Clearly the same general structure is present in all three figures. Note that the transverse JEF is smeared on a scale  $R$  compared to the underlying  $E_T$  distribution of Figure 1. For comparison the Snowmass cone jet algorithm[2, 7] identifies jets at a discrete set of values of locations  $m_j$  defined by the  $E_T$  weighted cone “stability” constraint. These stable cone locations  $m_j$  are the solutions of the equation

$$\int dm' E_T(m') \times (m' - m_j) \times A(m' - m_j) = 0. \quad (6)$$

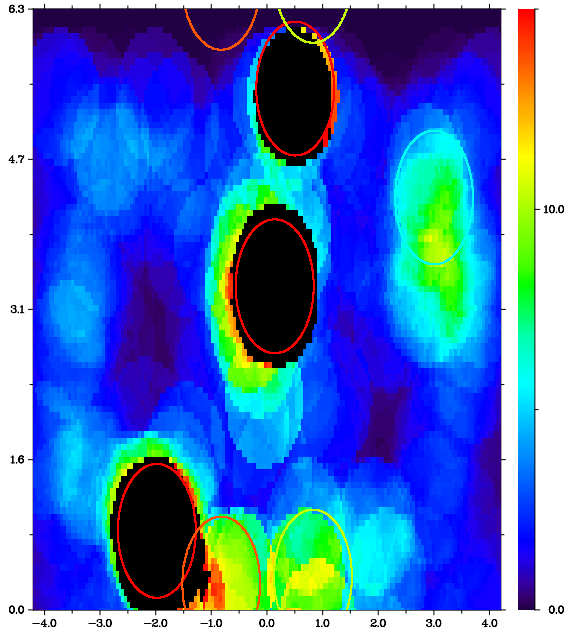


FIG. 3: Same as Figure 2 except that the maximum  $E_T$  for the color coding is 15 GeV.

The corresponding cone jet  $E_T$  values are found by evaluating Eq. 5 (times  $\pi R^2$ ) at the jet positions,  $E_{T,j} = \pi R^2 \times J_T(m_j)$ . The non-analytic character of the jet cone algorithm referred to earlier arises from the need to solve Eq. 6 and then work with only the discrete set of solutions, *i.e.*, the jets in an event.

### C. Observables

We can now proceed to define more general observables. The basic assumption of the JEF approach is that event-by-event each value of the direction  $m$  is equally likely to correspond to a jet with 4-momentum proportional to  $J_\mu(m)$ . Relative probabilities of observables having values in a specified range will correspond to the size of the area in  $m$  covered by the JEF with the correct range of values. To illustrate these ideas, consider a general  $n$ th order observable  $C_n$  represented by an  $n$ th order function of  $J_\mu(m)$ ,  $C(J(m_1), \dots, J(m_n))$ . The corresponding event probability distribution, including the possibility of providing a set of

angular cuts  $\Omega$ , is given by

$$P(C_n | \Omega(m_{cut})) = \left( \prod_{i=1}^n \int \frac{dm_i}{\pi R^2} \Omega(m_i - m_{cut}) \right) \delta(C_n - C(J(m_1), \dots, J(m_n))) \quad (7)$$

$$\propto \frac{d\sigma_{\text{JEF}}}{dC_n},$$

where we have normalized the area to the ‘‘cone size’’  $\pi R^2$ . (Note that one might also consider applying a cut directly on the  $J(m)$ , *e.g.*,  $J_0(m) > E_{cut}$ .) To determine the differential cross section for the observable  $C_n$  from an experiment we simply sum over events as

$$\mathcal{L} \frac{d\sigma}{dC_n} = \sum_{\text{events}} P(C_n | \Omega(m_{cut})), \quad (8)$$

where  $\mathcal{L}$  is the integrated luminosity. We obtain an event occupancy probability  $O$  by integrating over the probability function

$$O(C_n(\min), C_n(\max) | \Omega(m_{cut})) = \int_{C_n(\min)}^{C_n(\max)} dC_n P(C_n | \Omega(m_{cut})) \quad (9)$$

$$= \left( \prod_{i=1}^n \int \frac{dm_i}{\pi R^2} \Omega(m_i - m_{cut}) \right) \Theta(C_n(\max) - C(J(m_1), \dots, J(m_n)))$$

$$\times \Theta(C(J(m_1), \dots, J(m_n)) - C_n(\min)).$$

This final expression indicates that we are simply calculating the relative area in  $m$  for which the JEF has the correct value to yield the desired value of the observable. We can then count the number of events with effective occupancy number  $O$  and convert it into a cross section,

$$\mathcal{L} \sigma(C_n(\min), C_n(\max)) = \sum_{\text{events}} O(C_n(\min), C_n(\max) | \Omega(m_{cut})). \quad (10)$$

This formula can be used to construct bin values and the corresponding distribution. Let us illustrate these ideas by considering some explicit examples.

1. The JEF jet mass  $M(J(m)) = \pi R^2 \times \sqrt{J_\mu(m) J^\mu(m)}$  is an example of a  $C_1$  observable with a event probability distribution of the form

$$P(M_J) = \int \frac{dm}{\pi R^2} \delta(M_J - M(J(m))). \quad (11)$$



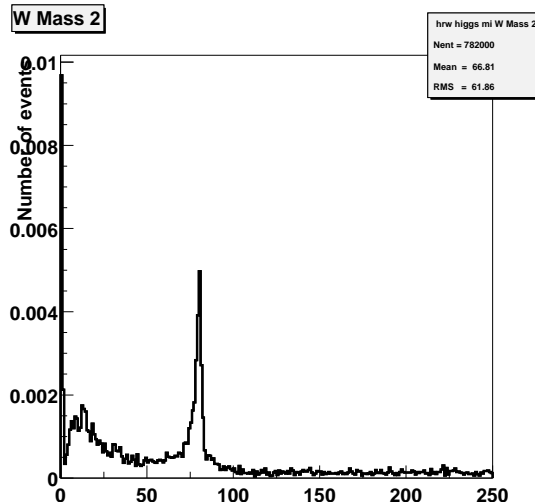


FIG. 4: The JEF jet invariant mass as defined in Eq. 11 for  $W$  boson decay in Higgs boson decay.

The corresponding occupancy probability has the form

$$O(M_J(\min), M_J(\max)) = \int \frac{dm}{\pi R^2} \Theta(M_J(\max) - M(J(m))) \Theta(M(J(m)) - M_J(\min))$$

Thus the fraction of the events with a JEF jet with mass in the specified range is proportional to the fractional area in the  $m$  plane occupied by JEF jets with a mass value in that range. To obtain the final distribution we sum over events. The (simulated) JEF jet mass distribution for the  $W$  decay into hadrons (treated as a single jet) in the process  $p\bar{p} \rightarrow H + X \rightarrow W^+W^- + X \rightarrow l\nu + \text{hadrons} + X$  is exhibited in Figure 4.

2. The JEF jet transverse energy  $E_T$  in the variables appropriate to a hadron collider is another  $C_1$  observable. The relative probability distribution for a CDF type rapidity acceptance and CDF  $E_T$  definition looks like

$$P(E_T | \Omega(0.1 < |\eta| < 0.7)) = \frac{1}{\pi R^2} \int_{0.1}^{0.7} d|\eta| \oint d\phi \delta(E_T - E(J(\eta, \phi)) \times \sin(\theta(\eta))) , \quad (12)$$

where the JEF energy distribution is given by  $E(J(\eta, \phi)) = \pi R^2 \times J_0(\eta, \phi)$ . As suggested above, we can obtain the effective occupancy number of JEF jets (per event) above an energy cut  $E_{T,\min}$  by integrating

$$O(E_{T,\min} | \Omega(0.1 < |\eta| < 0.7)) = \int_{E_{T,\min}} dE_T P(E_T | \Omega(0.1 < |\eta| < 0.7)) . \quad (13)$$

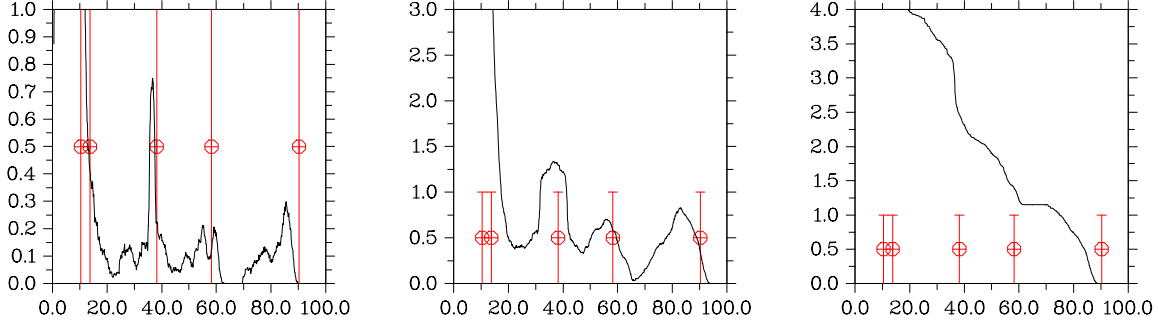


FIG. 5: The  $E_T$  probability function of Eq. 12 for the event of Figure 1: the left figure is for a bin width of 1 GeV and middle figure is for a bin width of 5 GeV. The right figure is the occupancy number of Eq. 13. The individual data points corresponding to the 5 largest energy jets found by the CDF cone jet algorithm and illustrated in Figure 1.

These quantities as evaluated for the sample jet event of Figure 1 are illustrated in Figure 5. The jets found by the standard CDF cone jet algorithm are also indicated as data points in the figures and correlate well with the peaks in the JEF probability distribution.

3. The JEF di-jet invariant mass  $M(J, J)$  is an example of a  $C_2$  observable with the form  $M^2(J(m_1), J(m_2)) = (\pi R^2)^2 \times (J(m_1) + J(m_2))_\mu (J(m_1) + J(m_2))^\mu$ , which assumes that the two JEF jets are non-overlapping.. The corresponding probability distribution is defined by

$$P(M_{JJ}^2) = \iint \frac{dm_1 dm_2}{(\pi R^2)^2} \delta(M_{JJ}^2 - M^2(J(m_1), J(m_2))). \quad (14)$$

#### D. An Example JEF Analysis

As an example of a JEF style analysis we briefly review a JEF di-jet analysis performed previously[8]. The goal is to calculate the differential transverse energy distribution of a jet in the CDF central rapidity strip,  $0.1 < |\eta| < 0.7$ , while requiring that a second jet with transverse energy at least as large as one half of the transverse energy of the central jet,  $E_{T,2} \geq E_{T,1}/2$ , is tagged in a forward region,  $1.2 < |\eta_2| < 1.6$ . The corresponding di-JEF

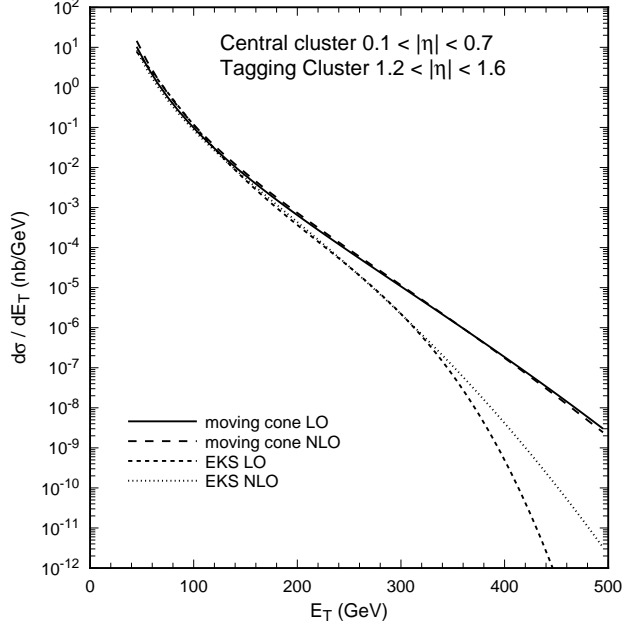


FIG. 6: Differential cross section of Eq. 16 for both JEF (moving cone) and EKS (fixed cone) analysis at LO and NLO.

probability density function for an event obeying the appropriate cuts is expressed as

$$\begin{aligned}
 P_{\text{di-jet}}(E_T) &= \frac{1}{(\pi R^2)^2} \iiint d\eta_1 d\phi_1 d\eta_2 d\phi_2 \Omega(0.1 < |\eta_1| < 0.7) \times \Omega(1.2 < |\eta_2| < 1.6) \\
 &\times \Omega(E_T(J(\eta_2, \phi_2)) > E_T(J(\eta_1, \phi_1))/2) \times \delta(E_T - E_T(J(\eta_1, \phi_1))).
 \end{aligned}
 \tag{15}$$

Using this probability function the desired differential cross section is obtained from Eq. 8

$$\frac{d\sigma_{\text{di-jet}}}{dE_T} = \frac{1}{\mathcal{L}} \sum_{\text{events}} P_{\text{di-jet}}(E_T).
 \tag{16}$$

The perturbative results for this cross section at LO and NLO for both the JEF analysis of Eq. 16 and using the standard cone jet analysis of EKS[7] are exhibited in Figure 6 for the case  $R = 0.7$ . These results illustrate some of the desirable features of the JEF approach. Note first that at smaller  $E_T$  values both jet definitions yield similar results at both LO and NLO. However, at larger  $E_T$  values, where the boundaries of phase space play are more relevant, the JEF result is larger than the “traditional” cone jet result and, more importantly, is less sensitive to the higher order corrections. We can understand this reduced sensitivity in terms of the smearing of the rapidity cuts in the JEF analysis. In the

JEF analysis the underlying partons can violate the rapidity cuts by as much as  $R$  and still contribute to a JEF style jet that respects the cut. For example, the parton contributing to the secondary jet is only required to obey  $|\eta_2| > 1.2 - R = 0.5$  in order to make a nonzero contribution. In contrast, the traditional cone jet with a single parton inside requires the parton to be collinear with the jet and thus in the current analysis receives contributions only from  $|\eta_2| > 1.2$ . This smearing of the details of the rapidity cuts explains both the larger magnitude and the reduced dependence on higher orders of the JEF di-jet analysis. We can expect a similarly reduced dependence on the stochastic effects of showering and hadronization.

### E. Concluding Remarks

We have discussed a different approach to the jet analysis of hadronic states, the JEF analysis, which follows from the earlier C-algebra formalism[5] and which differs from traditional algorithmic approaches in that unique jets are not identified event-by-event. Instead the analysis proceeds through the evaluation of Jet Energy Flow distributions. The brief discussion presented here suggests that the JEF style analysis of hadronic final states in hard scattering processes will provide observable measures of the underlying short-distance parton structure that are less sensitive to higher order corrections and to showering/hadronization corrections than more conventional jet algorithm analyses. Clearly much more needs to be done in order to demonstrate and make use of this conclusion.

- 
- [1] See, for example, the discussions in the Proceedings of recent Workshops: G. C. Blazey, *et al.*, *Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop*, hep-ex/0005012; P. Aurenche, *et al.*, *The QCD and Standard Model Working Group: Summary Report from Les Houches*, hep-ph/0005114.
  - [2] J. E. Huth, *et al.*, in *Proceedings of Research Directions for the Decade: Snowmass 1990*, July 1990, edited by E. L. Berger (World Scientific, Singapore, 1992) p. 134.
  - [3] This algorithm arose in the context of  $e^+e^-$  collisions. Its application to hadron colliders is discussed in S.D. Ellis and D. Soper, *Phys. Rev.* **D48**, 3160 (1993); S. Catani, Yu.L. Dokshitzer

- and B.R. Webber, Phys. Lett. **B285**, 291 (1992); S. Catani, Yu.L. Dokshitzer, M.H. Seymour and B.R. Webber, Nucl. Phys. **B406**, 187 (1993).
- [4] S.D. Ellis, J. Huston and M. Tönnesmann, *On Building Better Cone Jet Algorithms*, hep-ph/0111434.
- [5] F.V. Tkachov, Int. J. Mod. Phys. **A12**, 5411 (1997), hep-ph/9601308; F.V. Tkachov, Phys. Rev. Lett. **73**, 2405 (1994), hep-ph/9901332; F.V. Tkachov, *Jet algorithms. Wrapping up the subject*, hep-ph/0012210; D.Yu.Grigoriev and F.V.Tkachov, *An efficient implementation of the optimal jet definition*, hep-ph/9912415; F.V.Tkachov, *A theory of jet definition*, hep-ph/9901444.
- [6] Other recent applications of energy flow ideas include A.V. Belitsky, G.P. Korchemsky and G. Sterman, Phys. Lett, **B515**, 297 (2001) and C.F. Berger, T. Kucs and G. Sterman, *Energy Flow in Interjet Radiation*, hep-ph/0110004. The long term history of these ideas, especially in the context of  $e^+e^-$  collisions dates back, *e.g.*, to the work of G. Sterman and S. Weinberg, Phys. Rev. Lett. **39**, 1426 (1977) and G.C. Fox and S. Wolfram, Nuc. Phys. **B149**, 413 (1979). For an early application to hadronic collisions see R. D. Field, *Jets and Event Topologies at the SSC*, in the *Proceedings of the Snowmass Summer Study 1984*, ed. by R. Donaldson and J. G. Morfin, p. 713.
- [7] S.D. Ellis, Z. Kunszt and D. Soper, Phys. Rev Lett. **69**, 1496 (1992); Phys. Rev Lett. **64**, 2121 (1990); Phys. Rev. **D40**, 2188 (1989); Phys. Rev. Lett. **62**, 726 (1989).
- [8] W.T. Giele and E.W.N. Glover, *Probabilistic Jet Algorithms*, hep-ph/9712355.