Disclaimer

This note has not been internally reviewed by the DØ Collaboration. Results or plots contained in this note were only intended for internal documentation by the authors of the note and they are not approved as scientific results by either the authors or the DØ Collaboration. All approved scientific results of the DØ Collaboration have been published as internally reviewed Conference Notes or in peer reviewed journals.

DØ Note 2188
OPAL-TN 241
15 July 1994

Topological Quark/Gluon Jet Tagging

Paul Grannis

CERN

Abstract

We investigate a method for tagging quark and gluon jets produced in e^+e^- collisions based upon the flow of energy and multiplicity within a set of annular cones around a jet axis, and the correlations amongst these variables. The method gives good agreement with previously determined quark and gluon purity for a specific three-jet topology. Purities of about 70% can be achieved for quarks or gluons with efficiencies in the range of 55 - 60%.

0. Preface

This note was written to summarize work done using the data and Monte Carlo events from the OPAL detector now operating at LEP. It is an outgrowth of several OPAL analyses, published and in progress, to investigate the differences between quark and gluon jets. In e^+e^- collisions, the multihadron events are virtually all due to production of a $q\bar{q}$ pair from the intermediate Z/γ state formed in the primary collision. Subsequent gluon radiation from the quarks then gives rise to some three (or more) jet final states observed in the detector. The topology chosen for the OPAL analyses has a leading jet plus a pair of lower energy jets symmetrically displaced from the leading jet. By choosing the angle between the leading and lower jets to be large (about 150°), one has good assurance that the leading jet was one of the primary quarks. The lower energy jets so selected have an energy of about 25 GeV. Without further information, the two lower jets have about equal probability to be quark or gluon. In a subset of these events however, one can establish that one of the lower jets is likely to be a heavy quark by finding a non-zero separation of some tracks from the primary event vertex. In this case, the other low energy jet is most likely to be a gluon (about 80% purity). After deconvoluting the tagged and untagged samples to give statstically pure samples of quarks and gluons, OPAL has shown:

- The overall total (charged particles and EM clusters unassociated with tracks) multiplicity for gluon jets exceeds that for quarks by 1.27 ± 0.07. The excess in charged multiplicity in gluon jets is 1.33 ± 0.09. (These numbers pertain to the 'Durham' jet algorithm, but are similar for a cone algorithm.)
- The differential distribution of energy in the jet as a function of the angular separation from the jet axis is different for quarks and gluons, with the gluons showing more energy far from the axis. The peak density in $\mathrm{dE/d}\theta$ for quarks is at about 0.05 radians (θ here is the particle angle with respect to the jet axis), while for gluons it is at about 0.09 radians. When the energy distributions are normalized to unity, the quark distribution falls beneath the gluon distribution at about $\theta = 0.125$.
- The differential distributions of total multiplicity as a function of angular separation from the jet axis also differs for quarks and gluons, with the gluon distribution being broader. The peaks for the quark and gluon distributions occur at about $\theta = 0.07$ and 0.14 respectively. Beyond $\theta = 0.16$, the normalized gluon distribution exceeds the quark distribution.
- The number of particles with large energy fraction, x_E $(E_{particle}/E_{jet}) > 0.5$ is larger for quarks than for gluons.
- OPAL has begun a comparative study with CDF and DØ of these properties of high p_T jets; thus far, the qualitative conclusions are that jets are similar in the two environments, once the different quark/gluon contents are taken into account.

The analysis described in this note evaluates how well one might be able to distinguish quark from gluon jets on a jet-by-jet basis. Such tagging would be useful for some LEP physics studies, but is even more strongly motivated by $\bar{p}$-p physics. Often we have

the problem that an interesting signal which involves jets in the final state is swamped by background from uninteresting QCD processes. This is particularly the case when one seeks a signal that involves quark jets (*e.g.* from W decay). For these cases, DØ would very strongly benefit from a 'filter' which can be applied to preferentially select quarks – even if only a factor of three or so rejection of gluon jets can be achieved. A simple example would be the search for W production with subsequent hadronic decay. Quark tags would be of great use in reducing background in the top searches in which the W's decay hadronically (lepton + jets or all jets), and could well be powerful handles in identifying initial or final state radiation in the mass analyses of $t\bar{t}$ events. One can imagine a host of interesting applications in QCD physics – isolation of particular sub-processes in direct photon and $\bar{b}$ production, refined studies of gluon radiation patterns in multijet events, etc..

The work described here uses variables to characterize the jets taken from the previous OPAL analyses – the energy and multiplicity distributions within jet cones for charged particles and 'unassociated clusters' $\equiv$ EM calorimeter clusters. The OPAL EM calorimeter segmentation is approximately $(\Delta\eta = 0.04) \times (\Delta\phi = 0.04)$, roughly equivalent to the EM3 section of the DØ calorimeter. Of course, DØ has no charged particle momentum measurement, but instead has a finely divided hadron calorimeter whose energies can perhaps be substituted for the charged particle information. I would hope that an analysis similar to the one presented here could be undertaken in DØ, using perhaps an input variable set based on the pattern of energy deposits in the EM and Hadronic calorimeters.

The present work confines itself to nearly monoenergetic jets, so as to compare the tagging with the experimentally defined samples for gluons and quarks. Extension to higher energy jets would be necessary for a DØ application. However, the present results on jet energy profiles and the dependence on jet E_T seen in DØ suggests that the energy variation may not be too large.

1. Introduction

For a variety of QCD studies in e^+e^-collisions, it is of interest to distinguish quark from gluon jets. The multiplicity and energy flow differences between the two partonic species are themselves of interest and have been explored in previous OPAL publications, both for jets defined with the 'Durham' algorithm [1][2] and with a cone algorithm similar to those used in $\bar{p}$-p collider experiments [3]. These studies have quantified the differences between quark and gluon jets experimentally through the choice of two well-controlled data sets. Both data sets select the three-jet topology in which the second and third jets lie symmetrically within the interval $150^\circ \pm 10^\circ$ with respect to the leading jet. This topology strongly favors a quark assignment for the leading jet. In one data set, no further selection is made, so that the two lower energy jets are almost equally populated by quarks and gluons. In the second data set, one of the lower energy jets is tagged with a non-zero decay length using the silicon vertex detector and thus is dominantly (heavy) quark in origin. In this case, the other jet is enriched in gluon content. Deconvoluting the multiplicity and energy flow distributions in these two experimental samples to give pure gluon or quark data sets shows that the gluon total multiplicities exceed those for quarks by 1.27 ± 0.07 and that the angular distribution and energy flow of particles with respect to the jet axis is broader for gluons than for quarks. Monte Carlo simulations of quark and gluon jets are generally in good agreement with the experimental data. The strength of this analysis is its firm grounding in purely experimental information.

There are a variety of physics studies for which case-by-case identification of the parton parentage of jets is essential or useful. Among these are the studies of the density of particles emitted between two quark jets in comparison with that between quark and gluon, exploring the properties of the color string [4]; measurements of the QCD colour factors in multi-jet final states [5]; or studies of the relative content of specific particle types (*e.g.* protons, Λ's etc.) in quark or gluon jets. Measurements extracting the inclusive cross-sections for gluon emission in multi-hadron events would be of interest. Other possible extensions include the study of large x gluonic systems recoiling from a di-quark system [6], and systematic comparison of the radiation patterns in four or more jet final states with QCD predictions. The existence of tools for efficient tagging of quark/gluon jets of arbitrary energy or topology would enhance or enable all of these studies.

High statistics studies require that relatively high efficiency tags for gluons and/or quarks be available and that the purity of these tags be well understood. It is not necessary that the purity be 100%, since it is possible to extract information from samples with well-determined admixtures of quarks and gluons. The existing methods of tagging parton jets using detached vertices or semi-leptonic decays suffer from the relatively low tagging efficiency or from the biasses that these methods give for heavy-quark enriched quark jets. Methods based upon assigning the lowest energy jet as a gluon work well when the energy of the third jet is markedly below that of the two leading jets, but necessarily fails for symmetric partition of the energies.

A capability for tagging quark jets may also prove to be of some use in the LEP200 environment, in order to select events with W decays into di-quark final states without

gluon radiation. Filters for quark and gluon jets would be of great use in hadron collider experiments. There the problem is often the overwhelming preponderance of gluon jets from ordinary QCD processes, swamping the rarer but interesting signatures for states which decay into quarks. Examples include particularly the search for $t\bar{t}$ production, with its decay into two W's and two b jets. Insistence on the presence of quark jets in these cases could be a strong suppressant in the lepton + jets and all jets decay channels (where one or both W's decay into $q\bar{q}$), and would allow simplification of the final state jet assignments for measuring the top mass. Gluon identification would permit selection of the Compton graph for direct photon production, which is particularly sensitive to the gluon structure function, and would aid in disentangling the subprocesses responsible for $\bar{b}$ production. Finally, with a silver bullet to identify the gluons from the quarks, a variety of studies of QCD similar to those performed at LEP, but at substantially larger q^2, would become possible.

In this study, we investigate a tool for discriminating q and g using the topological properties measured in the previous OPAL publications. Such discriminants would be useful if their efficiency exceeded that of the silicon or lepton tagging techniques and have reasonably high purity ($\geq 2/3$?) of known accuracy. Previous attempts to perform such a separation have been reported [7], but they have two major limitations. First, they have been formulated in terms of cuts on a set of event shape variables, but have not included the possibility of correlations amongst these variables. Second, and more importantly, these studies have relied upon the properties of Monte Carlo generators to establish the distinction of the two partonic species. It is this second shortcoming which is obviated to some degree in OPAL with its sample of experimentally distinguishable samples of known gluon and quark content. Testing any differentiation scheme against these well-defined data sets is crucial in gaining confidence in the tagging methods.

The work reported here uses the same selection of data and Monte Carlo events as used in the previous OPAL studies [1][2][3] – namely events with exactly three coplanar jets in which the second and third jets are symmetrically displaced in angle with respect to the leading jet by $150° \pm 10°$. A discriminant for the partonic species is developed which utilizes the correlations amongst the kinematic variables chosen. This discriminant is studied for Monte Carlo jets whose quark/gluon character is known, and is then applied to the data sets of vertex-tagged gluon jets and to undifferentiated jets to show consistency with the previously established quark/gluon admixtures. Armed with reasonable consistency with data in this global sense, we proceed to calculate the jet-by-jet tagging efficiency and purity as a function of the cuts on the discriminant.

2. Covariance Matrix Discrimination

Given a set of variables which carry information about the parentage of a particular event (call the variable set $\{x_i\}$), a simple χ^2-like variable can be formed to represent the extent to which this event conforms to the properties of a specific parent sample. The method we employ here is the **H**-matrix formalism [8][9], in which a training sample (here the Monte Carlo quark or gluon events) is used to form a covariance matrix **M**:

$$M_{ij} = (1/N)\sum_{n=1}^{N}(x_i - \overline{x}_i)(x_j - \overline{x}_j) \quad , \qquad (1)$$

with $\overline{x}_i$ the mean values of the variables over the ensemble $\mathcal{E}$ of training events, and N the number of events in $\mathcal{E}$. Defining the inverse $\mathbf{H} \equiv \mathbf{M}^{-1}$, one can form a parameter, $\zeta^{(k)}$, which describes how well a particular new event, k, (not from the training sample) conforms to the ensemble $\mathcal{E}$:

$$\zeta^{(k)} = \sum_{i,j}(x_i^{(k)} - \overline{x}_i)H_{ij}(x_j^{(k)} - \overline{x}_j) \quad . \qquad (2)$$

The distribution of $\zeta^{(k)}$ conforms to the usual chi-square distribution for events originating from same physical source as $\mathcal{E}$.

In the present case, $\mathbf{H}$ can be constructed separately for quark and gluon jets from the Monte Carlo training sample and thus, for a particular event k, both ζ_q and ζ_g can be formed. The difference

$$\Delta\zeta = \zeta_g - \zeta_q \qquad (3)$$

is known as a Fisher discriminant and is a measure of the preference for this event to be assigned a gluon or quark parentage. Typically, $\Delta\zeta$ tends to be negative for gluons and positive for quarks.

It is interesting to note that there is a strong similarity between the **H**-matrix approach and the neural network algorithms. In the limit of large training samples on the same set of input variables, a correspondence can be shown for the hidden node weights in a neural network and the **H**-matrix elements [10]. The **H**-matrix method has some advantage in its similarity to the error-matrix correlation algebra commonly used in high energy physics and thus its relatively simpler interpretation. The method can also be used to predict the values of variables which are not measured directly in an experiment, but are correlated with those variables which are accessible (see Technical Note 232 for an application of this extension to the determination of the B-meson momentum in semileptonic decays in OPAL.)

3. Event Selection and Choice of Variables

In order to facilitate comparison of our results to the OPAL data [1][2][3], we have adopted an event selection which is identical to that used in previous analyses. Specifically, we used the analysis code of J.W. Gary in which event selection is based upon the following criteria:

- Exactly three jets are required, using a cone algorithm in $\theta - \phi$ space. The cone size is 0.7 radians and the minimum jet energy is 7 GeV. Charged tracks and unassociated clusters in the EM calorimeter (collectively 'particles') are used to define the content of the jet (energy and multiplicity).

- Jets are required to have a minimum of two particles; all jets must lie within $|\cos\theta| < 0.9$.

- The three jet system must be coplanar to within 2°.
- The center-of-mass energy is required to be at the Z peak.
- From these events, the subset satisfying the angular requirements that both lower energy jets have an angle with respect to the leading jet of between 140° and 160° is selected for final study. This corresponds to jet energies of about 25 GeV.

There are many choices of variables which could be chosen to describe and differentiate quark and gluon jets. The basic information expected to be useful for discrimination is the overall jet multiplicity and the distribution of energy and particle number flow as a function of angular distance from the jet cone. In addition, the distribution of particles in $x_E = (E_{particle}/E_{vis})$ has been shown to be different for gluons and quarks [6] (the x_E distribution for quarks is harder). The earlier CDF analysis [7] used as input variables the overall jet charged multiplicity, ratio of energy in the EM calorimeter to the total energy, and moments ($n = \pm 2$) of the momenta of charged particles in the jet rest frame:

$$K_n = \Sigma(k_{ti}/M)^n \quad ; \quad Q_n = \Sigma Q_i(k_{ti}/M)^n \quad ,$$

with M = the jet mass. Neutral clusters were not included in their moment variables. We feel that it is difficult to justify any specific choice of moments, but prefer to use variables which are closer to the measured quantities in a jet. Thus we adopt a choice of variables taken directly from the reconstruction of Monte Carlo or data events.

For a given jet (selected with the cone size of 0.7 radians) we define a set of ten annular subcones centered on the jet axis. The outer edges of these subcones are:

Cone edges, $\{\mathcal{R}_i\}$, = {0.02, 0.05, 0.08, 0.12, 0.18, 0.24, 0.30, 0.42, 0.54, 0.70},

where the specific choice is made to give sufficient granularity in the region $0 < \mathcal{R} < 0.3$ where the energy and multiplicity profiles for quark and gluon jets peak, and are most distinct from each other.

In the basic study, 22 possible event variables are recorded separately for gluon and quark jets; these are the total particle multiplicity, visible energy, and the energy and number of particles deposited within each of the ten subcones. The full set of variables, or any selected subset or mergings (*e.g.* energy sum for the $4th$ and $5th$ subcones), can then be chosen for construction of the **H**-matrices and $\Delta\zeta$ calculations.

In some studies, five additional variables were included, based on the observation that the x_E distributions of particles in quark jets differ from those in gluon jets. We chose these variables as the number of particles within the x_E ranges [0.1 - 0.2], [0.2 - 0.3], [0.3 - 0.5], [0.5 - 0.7], and [0.7 - 0.9]. Again any subset or merging of variables could be chosen subsequently.

4. Gluon/Quark Jet Discrimination

We formed Ntuples of these 22 (or 27) variables from analysis runs on SHIFT. For the Monte Carlo, about 1500 gluon and 1500 quark jets from the inclusive multihadron sample in Run 2218 were used. For the vertex-tagged data or untagged three jet data, DADlists from W. Gary were used to assure a common sample with the earlier analyses. We accumulated about 1200 tagged gluon (or quark) jets and about 5100 undifferentiated jets from the symmetric data sample The Ntuples served as input for generation of the **M** and **H**-matrices and subsequent calculation of $\Delta\zeta$.

The basic and most complete analysis was performed using the full set of 22 variables defined above (total multiplicity, E_{vis}, energy and multiplicity in the ten subcones). Subsequent comparisons were made (a) the case where both subcone energies and multiplicities in adjacent subcones were merged giving the coarser set of five subcones (12 variables), (b) for the case where the subcone multiplicity information was dropped (12 variables), and (c) the case where the multiplicities within the five x_E ranges were added to the basic 22 variable set.

Figure 1 shows the chi-square distributions ζ_g and ζ_q, calculated from the **H**-matrices from training samples of gluons and quarks respectively, for both true (Monte Carlo) gluons and quarks.. The basic 22 variable set was used for these calculations. There are small differences between quark and gluon hypotheses for either true gluons or quarks, but it takes a strong imagination to believe that these distributions by themselves offer any real discrimination power!

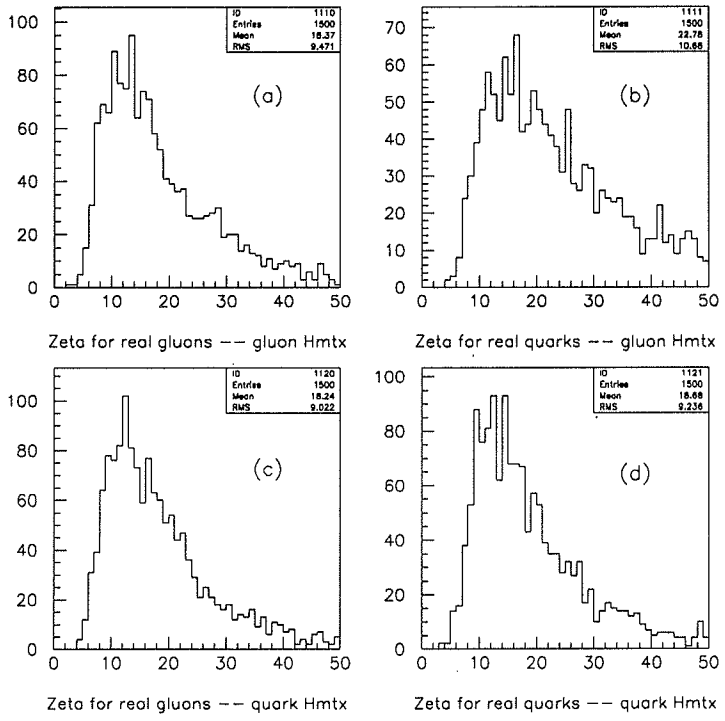

Figure 1. (a) Distribution of ζ_g ($\mathbf{H}_g$ used) for true Monte Carlo gluons; (b) Distribution of ζ_g for true quarks; (c) Distribution of ζ_q ($\mathbf{H}_q$ used) for true gluons; and (d) Distribution of ζ_q for true quarks.

The distributions of the Fisher discriminant, $\Delta\zeta$, are more revealing since they reflect the preference of any individual event to resemble gluon or quark patterns. The distribution of $\Delta\zeta$ for true Monte Carlo gluons and quarks is shown in Fig. 2. Though not fully separated, the $\Delta\zeta$ distributions for gluons and quarks are clearly different, with the gluons preferentially populating the negative, and quarks the positive, $\Delta\zeta$ regions.

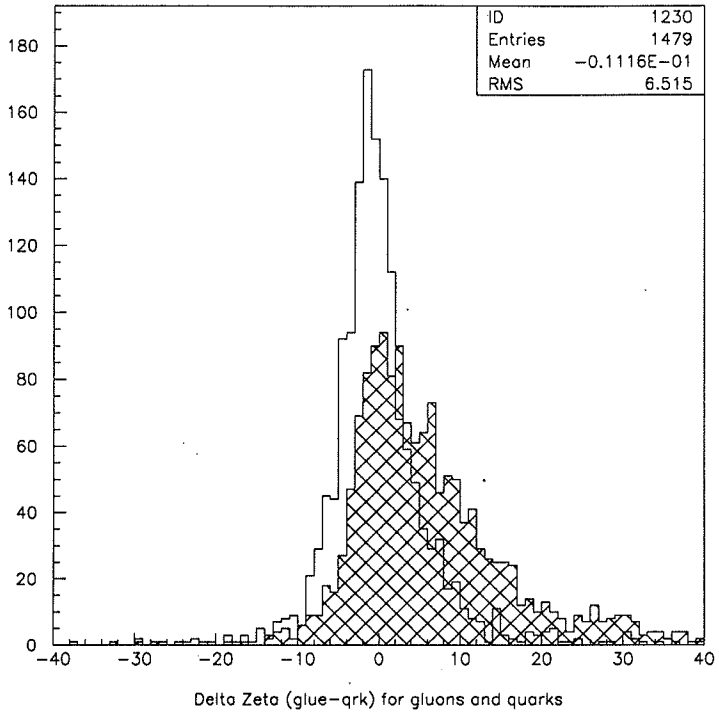

Figure 2. Distribution of $\Delta\zeta$ for true Monte Carlo gluons (unshaded) and for Monte Carlo quarks (shaded).

The OPAL data sample of tagged gluons was analyzed using the **H**-matrices determinded from the Monte Carlo samples. The $\Delta\zeta$ distribution for the tagged gluon jets, and for the tagged quark jets showing the separated vertex (dominantly heavy quarks), are shown in Fig. 3. Also shown in Fig. 3 is the $\Delta\zeta$ distribution for data jets taken from the sample in which no vertex tagging was performed, and for which gluons and quarks are expected to contribute approximately equally.

To give a check that our characterizations of jets are sensible, we can analyze a particular subject sample (*e.g.* tagged data gluons) in terms of their gluon/quark contents and compare these with what is known from independent sources. To make this decomposition, we perform a binned log-likelihood maximization of a particular subject $\Delta\zeta$ distribution to the sum of the $\Delta\zeta$-distributions for Monte Carlo gluons and quarks (Fig. 2), with weights f_g and f_q ($f_g + f_q = 1$). Table 1 shows the result of these decompositions from the observed $\Delta\zeta$ distributions, together with the 'expected' fraction. The 'expected' fraction is just the known composition by construction in the case of Monte Carlo samples; for the data samples it is the number taken from Ref. [1].

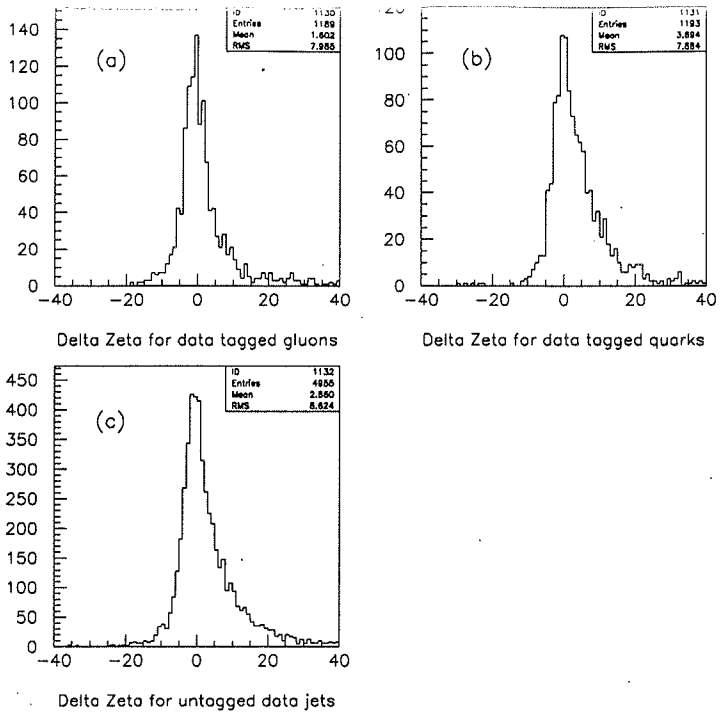

Figure 3. (a) Distribution of $\Delta\zeta$ for tagged data gluon jets; (b) for tagged data quark jets; and (c) for untagged data (quark and gluon) jets.

Table 1 Fractional gluon and quark content for various data samples.

Subject Sample	f_g	f_q	Expected f_g
100% Monte Carlo gluons	1.06 ± .09	-0.06 ± .07	1.00
100% Monte Carlo quarks	-0.11 ± .08	1.11 ± .10	0.00
50% MC gluons/50% quarks	0.52 ± .04	0.48 ± .04	0.50
Data tagged gluon jets	0.76 ± .06	0.24 ± .05	0.80
Data tagged quark jets	0.30 ± .07	0.70 ± .07	*
Data symmetric jets	0.52 ± .03	0.48 ± .03	0.48

The agreement between our determination for the symmetric jet data and the gluon tagged sample is reasonably good (of order 5%). We note that although we have determined the relative fraction of quark and gluon jets in the 'tagged quark' sample, this test is not easy to interpret. The **H**-matrices for quarks were performed with a flavor-averaged sample from the Monte Carlo with no requirement for vertex-tagging, whereas these data jets are vertex-tagged, and thus dominantly due to b's and c's.

We conclude that the agreement of our gluon and quark composition fractions and the known values is acceptable, and that at least on average, our **H**-matrix/$\Delta\zeta$ formalism

gives a good representation of the gluon (quark) content of known data samples. Thus we proceed to attempt discrimination at the individual jet level.

The distribution of $\Delta\zeta$ values shown in Fig. 2 gives some basis for jet-by-jet differentiation of gluon- and quark-induced jets. Clearly the distinction will not be perfect since the two species have overlapping $\Delta\zeta$ distributions. However, by establishing a cut ($\Delta\zeta_{LO}$) somewhere in the vicinity of $\Delta\zeta$ = (-2.0 to 0.0) and keeping only jets whose $\Delta\zeta$ lies below this cut value gives an enriched sample of gluons. Similarly with a cut ($\Delta\zeta_{HI}$) in the vicinity of $\Delta\zeta$ = (2.0 to 4.0) and keeping only events above this cut gives primarily quarks. For any given values of $\Delta\zeta_{LO}$ and $\Delta\zeta_{HI}$, we define a gluon candidate sample $\mathcal{G}$ as the set of events for which $\Delta\zeta \leq \Delta\zeta_{LO}$, and a quark candidate sample $\mathcal{Q}$ as the set of events for which $\Delta\zeta \geq \Delta\zeta_{HI}$. To the extent that $\Delta\zeta_{LO} \neq \Delta\zeta_{HI}$, there is an inefficiency due to the jets which lie within the range $\Delta\zeta_{LO} \leq \Delta\zeta \leq \Delta\zeta_{HI}$, but this inefficency need not be large and keeping $\Delta\zeta_{LO} < \Delta\zeta_{HI}$ permits us to form the samples $\mathcal{G}$ and $\mathcal{Q}$ with reasonable purities.

To make quantitative estimates for purity and efficiency as a function of $\Delta\zeta_{LO}$ and $\Delta\zeta_{HI}$, we compute the numbers of true gluon and quark jets (in the Monte Carlo samples) in the regions $\mathcal{G}$ and $\mathcal{Q}$: $N_{\mathcal{G}}^{gl}$ is the number of true gluon jets in the region $\mathcal{G}$, $N_{\mathcal{G}}^{qk}$ is the number of true quark jets in the region $\mathcal{G}$, $N_{\mathcal{Q}}^{gl}$ is the number of true gluon jets in the region $\mathcal{Q}$, and $N_{\mathcal{Q}}^{qk}$ is the number of true quark jets in the region $\mathcal{Q}$. Then we define purities $\mathcal{P}_g$ and $\mathcal{P}_q$:

$$P_g = N_{\mathcal{G}}^{gl}/(N_{\mathcal{G}}^{gl} + N_{\mathcal{G}}^{qk}) \quad ; \quad P_q = N_{\mathcal{Q}}^{qk}/(N_{\mathcal{Q}}^{qk} + N_{\mathcal{Q}}^{gl}) \quad ,$$

and efficiencies $\mathcal{E}_g$ and $\mathcal{E}_q$:

$$\mathcal{E}_g = N_{\mathcal{G}}^{gl}/N_{TOT}^{gl} \quad ; \quad \mathcal{E}_q = N_{\mathcal{Q}}^{qk}/N_{TOT}^{qk} \quad ,$$

where N_{TOT}^{gl} and N_{TOT}^{qk} are the total numbers of true gluons and quarks in the full $\Delta\zeta$ plot.

Contours of $\mathcal{P}_g$ and $\mathcal{P}_q$ as a function of $\Delta\zeta_{LO}$ and and $\Delta\zeta_{HI}$ are shown in Fig. 4. We do not consider the case $\Delta\zeta_{LO} > \Delta\zeta_{HI}$. We see that the purities tend to increase as the $\Delta\zeta_{LO}$, $\Delta\zeta_{HI}$ values move away from the boundary at $\Delta\zeta_{LO} = \Delta\zeta_{HI}$. The contours of $\mathcal{E}_g$ and $\mathcal{E}_q$ are shown in Fig. 5. Efficiencies fall as $\Delta\zeta_{LO}$ or $\Delta\zeta_{HI}$ diverge from the boundary at $\Delta\zeta_{LO} = \Delta\zeta_{HI}$. Thus there is an optimization to be performed to find a satisfactory choice of $\Delta\zeta_{LO}$ and $\Delta\zeta_{HI}$ cuts. In general this choice would depend upon the application one had in mind; for tagging jets with high efficiency but moderate purity, one would stay close to the $\Delta\zeta_{LO} = \Delta\zeta_{HI}$ boundary. If high purity is desired, one would operate far from the boundary at the sacrifice of efficiency. Since there is no unique criterion for choosing the cuts, we simply record some typical values in Table 2.

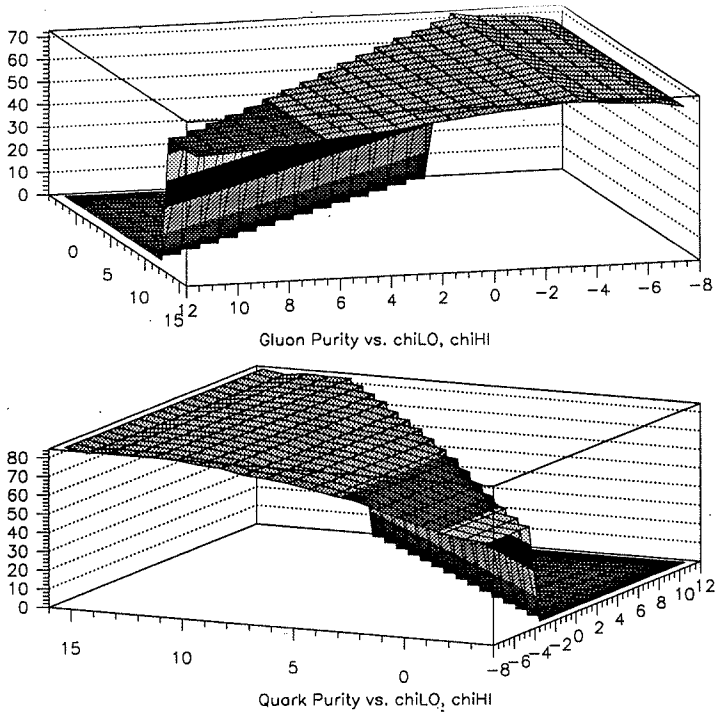

Figure 4. Contour plots of (a) gluon purity vs. $\Delta\zeta_{LO}$ and $\Delta\zeta_{HI}$ and (b) quark purity. The $\Delta\zeta_{LO}$ axis runs from -8. to 12. and the $\Delta\zeta_{HI}$ axis from -4. to 16.. No purity calculation is done for $\Delta\zeta_{LO} > \Delta\zeta_{HI}$.

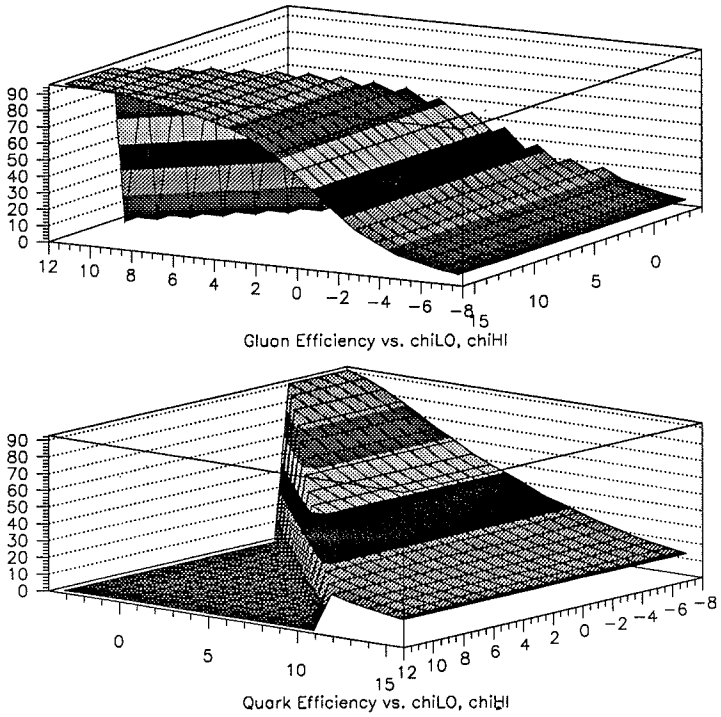

Figure 5. Contour plots of (a) gluon efficiency vs. $\Delta\zeta_{LO}$ and $\Delta\zeta_{HI}$ and (b) for quark efficiency. The $\Delta\zeta_{LO}$ axis runs from -8. to 12. and the $\Delta\zeta_{HI}$ axis from -4. to 16.. No efficiency calculation is done for $\Delta\zeta_{LO} > \Delta\zeta_{HI}$.

Table 2 Purity and Efficiency for gluons and quarks
for standard 22 variable **H**-matrix; (coarser 5 subcone binning);
and [10 subcones of energy but not multiplicity information.]

$\Delta\zeta_{LO}$	$\Delta\zeta_{HI}$	$\mathcal{P}_g$	$\mathcal{P}_q$	$\mathcal{E}_g$	$\mathcal{E}_q$
-4.	-2.	.73 (.74) [.59]	.56 (.54) [.52]	.19 (.08) [.06]	.85 (.93) [.90]
	0.	.73 (.74) [.59]	.62 (.62) [.63]	.19 (.08) [.06]	.73 (.76) [.73]
	2.	.73 (.74) [.59]	.69 (.72) [.70]	.19 (.08) [.06]	.61 (.54) [.56]
-2.	0.	.70 (.75) [.67]	.62 (.62) [.63]	.35 (.21) [.20]	.73 (.76) [.73]
	2.	.70 (.75) [.67]	.69 (.72) [.70]	.35 (.21) [.20]	.61 (.54) [.56]
	4.	.70 (.75) [.67]	.74 (.78) [.75]	.35 (.21) [.20]	.50 (.39) [.44]
0.	2.	.68 (.69) [.68]	.69 (.72) [.70]	.57 (.55) [.58]	.61 (.54) [.56]
	4.	.68 (.69) [.68]	.74 (.78) [.75]	.57 (.55) [.58]	.50 (.39) [.44]
	6.	.68 (.69) [.68]	.77 (.81) [.78]	.57 (.55) [.58]	.41 (.27) [.33]
2.	4.	.66 (.64) [.64]	.74 (.78) [.75]	.74 (.80) [.77]	.50 (.39) [.44]
	6.	.66 (.64) [.64]	.62 (.62) [.63]	.74 (.80) [.77]	.41 (.27) [.33]
	8.	.66 (.64) [.64]	.69 (.72) [.70]	.74 (.80) [.77]	.33 (.19) [.26]

In each column of Table 2, the first entry in the purity or efficiency is computed with the **H**-matrices based upon the standard choice of 22 variables. We have investigated the sensitivity to the basic choice of variables characterizing the jets in several ways. First we have coarsened the binning of the annular subcones defined for computing the energy and multiplicity flow in the jets. The purities and efficiencies for the case with five instead of ten subcones (merging cones (1 and 2), (3 and 4), (5 and 6), (7 and 8), (9 and 10)) are shown as the second entries in Table 2 (the numbers in parenthesis). The second case is constructed by dropping the subcone multiplicity information altogether, but retaining ten subcones for the energy flow. These purities and efficiencies are shown in Table 2 as the third entries [in brackets]. We see that coarsening the subcone definition or dropping the multiplicity flow information does degrade the tagging performance somewhat, but the effect is not drastic. The most powerful information for quark/gluon discrimination appears to be the energy flow distributions.

Following the observation that the particle energies tend to be higher for quark jets [1] [2], we extended the calulculation to include five additional variables (the number of particles observed in five ranges of $\boldsymbol{x}_E$) as described in Section 3. The $\Delta\zeta$ distributions using the expanded 27-variable **H**-matrix is very similar to that shown in Fig. 2 for the 22 variables. Comparison of the purities and efficiencies for the two cases (both with ten

subcones) is shown in Table 3; the entries in parentheses are for the 27 variable case. No substantive improvement is found using the extra information from the x_E-distributions, presumably because only a small fraction of jets contain particles in the high x_E region where quarks and gluons are different.

As an example of defensible cuts for the basic 22 variable calculation, setting $\Delta\zeta_{LO} = 0.0$ yields a gluon purity of $\mathcal{P}_g = 0.68$ for an efficiency $\mathcal{E}_g = 0.57$. Setting $\Delta\zeta_{HI} = 2.0$ gives a quark purity $\mathcal{P}_q = 0.69$ and efficiency $\mathcal{E}_q = 0.61$. Some improvement in purity can be acheieved by lowering $\Delta\zeta_{LO}$ or raising $\Delta\zeta_{HI}$, but as seen in Table 2 the efficiency falls quite rapidly as one departs from $\Delta\zeta_{LO}$= $\Delta\zeta_{HI}$ while the gains in purity are modest.

Table 3 Purity and Efficiency for gluons and quarks
for standard 22 variable **H**-matrix and (expanded 27 variable set).

$\Delta\zeta_{LO}$	$\Delta\zeta_{HI}$	$\mathcal{P}_g$	$\mathcal{P}_q$	$\mathcal{E}_g$	$\mathcal{E}_q$
-2.	0.	.70 (.75)	.62 (.60)	.35 (.36)	.73 (.76)
	2.	.70 (.75)	.69 (.65)	.35 (.36)	.61 (.63)
	4.	.70 (.75)	.74 (.69)	.35 (.36)	.50 (.52)
0.	2.	.68 (.71)	.69 (.65)	.57 (.54)	.61 (.63)
	4.	.68 (.71)	.74 (.69)	.57 (.54)	.50 (.52)
	6.	.68 (.71)	.77 (.71)	.57 (.54)	.41 (.44)

4. Conclusions

Using a method which incorporates the correlations among variables observed in jets, we have constructed a formalism for tagging (flavor independent) quark or gluon jets. The variables chosen are the energy and multiplicity flow within a set of ten subcones centered on the observed jet axis found in a cone algorithm. The calculation was performed for a specific topology of jets (from the symmetric three-jet topology with the two lower energy jets at $150° \pm 10°$), so as to facilitate comparison with the OPAL data set using the same selection criteria, with and without secondary jet tagging. On average, the quark and gluon contents of the data jets derived from the measure introduced here agree well with those deduced from the prior study [1]. To within the precision of this agreement (about 5%), we conclude that the JETSET Monte Carlo used to model the jet fragmentation is adequate.

Applied to individual quark and gluon jets in the Monte Carlo sample, the algorithm yields larger efficiencies for tagging quarks or gluons than is available from the observation of displaced vertices or leptons from quark semileptonic decays. The selection of quark jets

is not biassed toward heavy flavor species. Although we studied a particular energy of jets (due to the topology selection), there should be no difficulty in extending the method to arbitrary energy jets. Typical values for the purity and efficiency for tagging gluon or quark jets are 70% and 55 - 60%. The values of purity and efficiency can be tuned by varying cut parameters on the one-dimensional distributions of the discriminant variable, $\Delta\zeta$, trading purity for efficiency for quarks and gluons independently. The results of the method are not particularly sensitive to the specific subcone sizes. The success of the discrimination appears to rely primarily on the energy flow within the subcones; multiplicity flow and the particle x_E distributions appear to have less effect on the tagging performance.

We envision that measurements which require high statistics samples of enriched quark or gluon content will benefit from this tagging technique, particularly in those cases where a high precision determination of tagging purity is not required.

References

1. OPAL Collaboration, P.D. Acton *et al.*, Z. Phys. **C58** (1993) 387.
2. OPAL Collaboration, G. Alexander *et al.*, Phys. Lett. **B265** (1991) 462.
3. J.W. Gary, March OPAL Plenary Week, QCD Meeting. For related work using a cone algorithm and comparisons with hadron collider data, see OPAL Collaboration, R. Akers *et al.*, CERN-PPE/94-51, submitted to Z. Phys. **C**.
4. OPAL Collaboration, M.Z. Akrawy *et al.*, Phys. Lett. **B261** (1991) 334.
5. OPAL Collaboration, Physics Note 133.
6. J.W. Gary, Phys. Rev. **D49** (1994) 4503.
7. CDF Collaboration, "Jet Structure and Quark/Gluon Separation in CDF", M. Ninomiya, Proceedings of the 9th $\overline{p}$-p Workshop, Tsukuba, Japan (1993).
8. R. Engelmann *et al.*, Nucl. Inst. and Meth. bf 216, 45 (1983).
9. M. Narain (DØ Collaboration), "Proceedings of the American Physical Society Division of Particles and Fields Conference", Fermilab (1992), ed. R. Raja and J. Yoh.
10. H. Prosper, "Some Mathematical Comments on Feed-Forward Neural Networks", DØNote 1606 (1993) (unpublished).