SLAC–PUB–8360
January 2000

RECONSTRUCTION OF HIGH MASS PARTICLES FROM HADRONIC JETS AT A HIGH ENERGY LEPTON COLLIDER*

Gary R. Bower

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309

Abstract

Identifying the parent high mass particles that produce hadronic jets at future high energy lepton colliders will be a new challenge. We present here initial results of a study of jet discriminators used to identify the jets in the three processes $\gamma/Z \to qq$,ZZ and tt. We conclude that with good detector resolution it will be possible to reconstruct the parent particles reasonably well.

Talk presented at the International Workshop on Linear Colliders
Sitges, Barcelona, Spain
April 28 - May 5, 1999

*Work supported by Department of Energy contract DE–AC03–76SF00515.

1 Introduction

A new era in e^+e^- collider physics was reached in 1997 when LEP II beam energies reached the W mass. For more than 20 years since the first colliders in the 70's, all events with jets had their origin in the process $\gamma/Z \to qq$. As we contemplate the 1 TeV energy range proposed for the next generation of lepton colliders, the number of processes that produce multiple high mass states with hadronic fragmentation (jet) decay modes increases rapidly. Thus, we are faced with the new task of sorting out the origin of the jets we will find in these events. (A somewhat similar situation exists for the Tevatron and LHC and perhaps some of the ideas presented here will be applicable in that environment as well.)

By eye we will simply see events with several jets. To identify their origin we will have to measure properties of these jets and their interrelations. To select discriminators we will have to consider how the properties of the different underlying high mass state particles will express themselves in the final state jets.

This is a progress report on identifying jet discriminators and measuring their efficacy. The bulk of the work presented discusses several discriminators using generator level simulated data. Of course, it will ultimately be the job of an imperfect detector to provide the measurements and a further aim here is to study the effectiveness of different detector designs to measure jet properties. At the end of the report is a very preliminary and limited look at data from simulated detectors.

2 Description of the study

The number of high mass particles that have been proposed as theoretical possibilities to search for in the 1 TeV center of mass range is quite large. To make a comparative study of the jet properties of all these states initially is a case of too many trees obscuring the view of the forest. Instead we will try to achieve some general orientation towards the problem by focusing on discriminating between three processes that have significant interesting differences and which we already know will be produced copiously.

These processes are $\gamma/Z \to qq$, ZZ, and tt where q is a light quark (u,d,s,c or b) and t is the top quark. The Z and t are high mass particles which contrasts with the q case. The qq and tt states will have color connections between the quark pair but the Zs in ZZ do not. For each discriminator that is discussed, these three processes will be compared. In most of the figures which follow there will be three histograms in a vertical stack. The top histogram will

always be the distribution of the qq sample, the middle histogram will be the ZZ sample and the bottom histogram will be the tt sample. The PYTHIA 5.7 [1] generator was used in the study.

These studies should be carried out at a variety of energies but this initial investigation is made with beam energies of 250 GeV. Because of beamstrahlung and initial state radiation, the actual energy of interaction is often less than 500 GeV. All events are boosted to their center-of-mass for the determination of their discriminator values and the plots presented exclude events with a boost greater than 0.2c. For the generator level studies all charged and neutral final state particles including neutrinos are used. Events are required to have at least 5 particles in each hemisphere to eliminate $Z \to ee$, $\mu\mu$, and $\tau\tau$ where one of the τs decays leptonically.

3 Discriminators

3.1 Jet Invariant Mass

The most obvious distinguishing feature of jet-producing particles is their mass. Thus, perfect reconstruction of parent particle mass would be sufficient for identification; however, this is not possible for many reasons. The most serious difficulty is correctly identifying the measured final state particles that were produced by a given parent particle. It is easy to misassign a final state particle to a jet where it does not belong. This is true even at the generator level where we may assume that we can find, correctly identify, and perfectly measure all particles in the event including neutrinos. Of course, in the case of color-connected jets the concept of "particle belonging to a jet" is not even well defined.

To determine the parent particle mass in qq, ZZ, and tt events the most successful approach was to determine the event thrust axis and calculate the invariant mass of each hemisphere. The results are presented in Figure 1.

Two things are clear from the histograms. The distributions peak where they should but the tails are wide and overlap. Thus, the hemisphere mass can only provide a probablistic estimate of the parent particle. All of the discriminators studied show a similar pattern of probabilistic but indefinite identification. However, using several discriminators in conjunction can yield good discriminating power as will be seen below.

3.2 Number of jets

The nature of q, Z and t fragmentation/decay is that they are expected to produce one, two and three jets respectively disregarding gluon radiation and

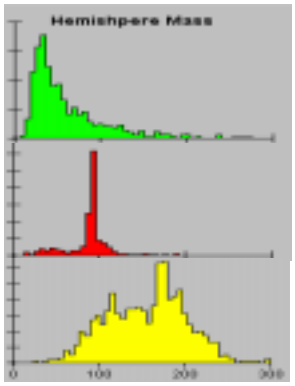

Figure 1: Hemisphere mass in GeV.

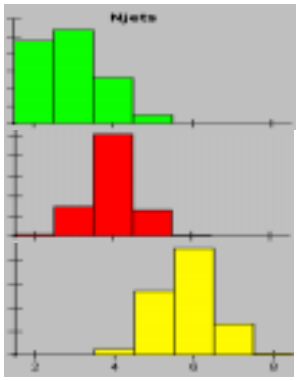

Figure 2: Number of jets per event.

other effects. Thus, the number of jets produced should be a helpful discriminator. Identifiying jets requires the use of a jet finding algorithm and selection of a y_c value to specify when the algorithm should stop combining particles into a single jet. The Durham[2] algorithm was used with a y_c of 0.5.

The results of measuring the number of jets in qq, ZZ and tt events are shown in Figure 2. As with invariant mass it is apparent that the peaks are where they should be, but the tails overlap and the number of jets is again useful only as a probabilistic discriminator.

3.3 Final state particle rapidity

The rapidity of the particles in these events are influenced by two factors. All other things being equal (they are not) the higher the mass of the jet initiating particle the lower the average rapidity of the final state particles. However, working in a different direction the color singlet states should have higher rapidity than the color connected states. These two trends are apparent in Figure 3 which shows the result of simply histogramming the rapidity of every final state particle in every event in the samples used.

The mass variable histogrammed in Figure 1 is a characteristic of a hemisphere and the number of jets variable in Figure 2 is a property of the event as a whole, whereas the variable in Figure 3 is a property of individual particles in the events and therefore is not directly a jet discriminator. Figure 3 shows distributions which are characteristic of the different processes. The rapidity distribution for each process can be characterized by variables such as their moments which can function as discriminators.

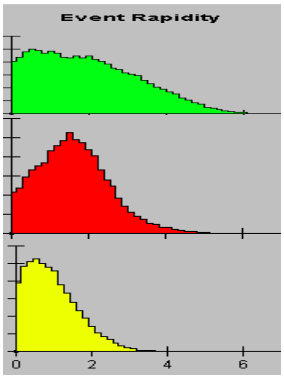

Figure 3: Rapidity of all final state particles in all events.

An obvious discriminating feature of the distributions in Figure 3 is their width. Figure 4 is a histogram of the rms width of the rapidity distribution of each event. Figure 4 shows that the width provides some discriminating power. Similarly, Figure 5 shows that the mean of the rapidity distribution of each event has discriminating power.

Two additional rapidity related variables which have some discriminating power are shown in Figures 6 and 7. The particle rapidity distribution in Figure 3 shows that essentially all the particles in tt events have rapidity less than 3 but the ZZ and qq events have some percentage of their particles with a rapidity greater than 3. Thus, for each event we measure the percentage of the total number of particles in the event with rapidity greater than 3 and histogram that variable in Figure 6. We see a discriminator that is very efficient at both identifying tt and eliminating qq. Figure 7 histograms a similar quantity, the percentage of the total number of particles in each event with rapidity less than 1. This variable gives good separation of tt from qq and ZZ events.

3.4 Event shape variables

The differences in mass, number of jets, and color connectedness of the three event types lead to differences in the shape of the events. A qq event with no gluon radiation will produce a pencil shaped event whereas a tt event will produce a much more spherically shaped event. The lack of color connection between the Z's in the ZZ events means that each Z will decay independently into a pencil shaped qq "subevent". Thus, in each Z hemisphere we expect

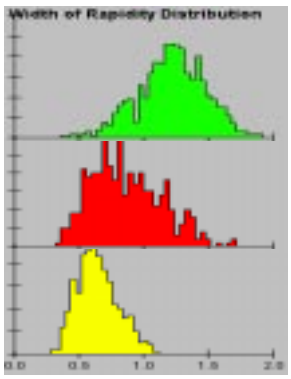

Figure 4: Event rapidity distribution rms width.

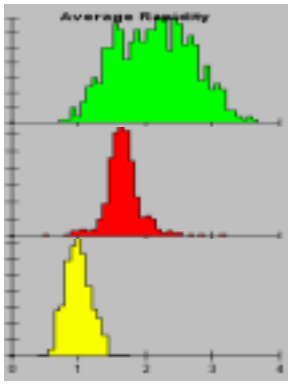

Figure 5: Event average rapidity.

to see a pencil shape but in the event as a whole, combining both Z decays we should see more of a spherical shape. We quantify these ideas with the following variables. What is presented is just a sample of what is possible.

The concept of energy flow in an event is formalized in the definition of the thrust tensor [3] which identifies principal axes called event, major and minor and a thurst value between 0 and 1 along each axis.

Consider the ratio of the event thrust to the major thrust of an event. For a pencil shaped qq event, the energy is predominantly along the event thrust axis and thus the ratio will be a number generally much larger than 1. On the other extreme a very spherical tt event will have a ratio close to 1. These expectations are evident in Figure 8 where this ratio is plotted and the great variation between the three event types is evident.

The oblateness of an event is defined as the major thrust less the minor thrust. We may also apply these event shape measures to a subset of the particles in an event. In particular it is interesting to consider their measures when applied to the particles in each hemisphere defined by the event thrust axis. Figure 9 shows a plot of the oblateness of each hemisphere. This kind of discriminator is particularly interesting because each event is represented twice. Although not demonstrated here in a plot we note that requiring an event to have at least one hemisphere with oblateness greater than 0.2 discriminates heavily against tt events.

As a final example of an interesting event shape variable, consider a ZZ event thrust axis and then consider boosting to the rest frame of each Z. In this frame the decay of the Z into a qq pair will have its own (hemisphere)

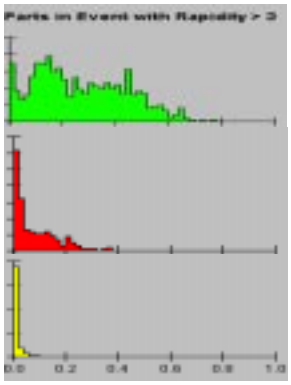

Figure 6: Fraction of total number of particles in an event with rapidity > 3.

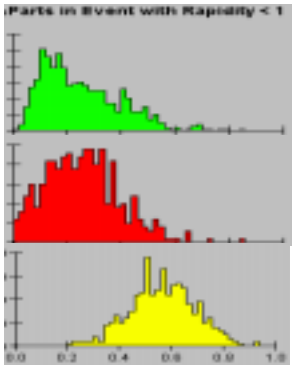

Figure 7: Fraction of total number of particles in an event with rapidity < 1.

thrust axis. This hemisphere thrust axis can be at any angle, θ, with respect to the overall event thrust axis. Now repeat this same gedanken experiment with a qq event. Due to the color connection when we boost to the hemisphere rest frame we expect the hemisphere thrust axis to align with the overall event axis. Figure 10, where we plot the boosted hemisphere thrust axis $cos\theta$, confirms these expectations. Depending on what one wishes to discriminate for or against, requiring conditions on the values in both hemispheres of an event is quite powerful. For example, requiring the absolute value of both to be close to 1 separates qq very nicely from ZZ.

One further interesting observation demonstrates the kind of subtle analysis possible to understand and identify the structure of events with jets. The mass plot in Figure 1 for ZZ events is bimodal with a small secondary peak at about half the Z mass. Figure 11 shows the ZZ mass plot overlaid with the mass of ZZ event hemispheres which have their hemisphere thrust axis parallel to the event thrust axis, *ie* with $|cos\theta| > 0.95$.

The probable interpretation of this is a Z with it's thrust axis along the event axis has a very high momentum particle pointing back towards the origin which is incorrectly assigned to the other hemisphere and thus to the other Z. Because the lost particle has high energy, the calculated mass of the true parent Z will be diminished. However, the momentum of the particle as seen from the misassigned parent will be lower due to the boost in the other direction, therefore having limited effect on the mass calculated for the false parent Z. Hence, there is a low mass bump but no high mass bump. It may be possible to identify the misassigned particles and correctly assign them.

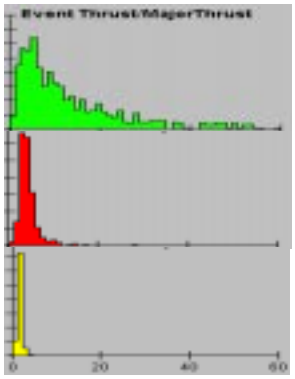

Figure 8: Ratio of event thrust to major thrust.

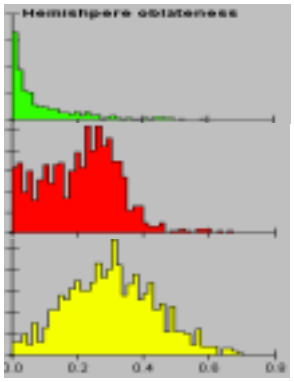

Figure 9: Hemisphere oblateness.

4 Identifying events

Since there are several useful discriminators, some method of combining them into an identification scheme is required. The goal of this study is to develop methods for identifying the individual parent particles in more complex processes like $\gamma/Z \to ZZH$ but for the present we will focus on simply identifying the three basic event types. The discriminators appear to lend themselves nicely to a neural net approach but this study has not yet progressed to that point either.

Instead, a brief effort was made to select sets of cuts that would be efficient at identifying qq, ZZ and tt events. Also, cuts were identified to decrease the background contamination of the samples of ZZ and qq events by events of the other two types. In the tables that follow each row describes the result of a set of cuts applied to each of the three event samples indicated in the column. Each box contains the percentage of the events in the sample designated by the column heading which passed the cuts. The first three rows are the results of the cuts for efficiency for qq, ZZ, and tt events. The last two rows are the results of tightening the efficiency cuts for qq and ZZ to eliminate contamination from the other two event types.

It should be emphasized that with a serious effort at tuning these cuts they should improve and it seems likely that a more sophisticated approach, such as a neural net, should perform even better.

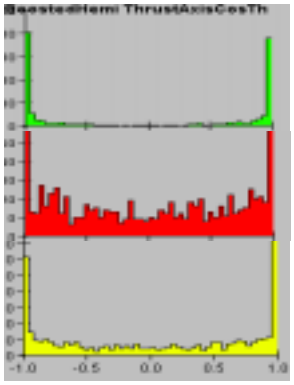

Figure 10: Cosine of the angle between the event thrust axis and the hemisphere thrust axis.

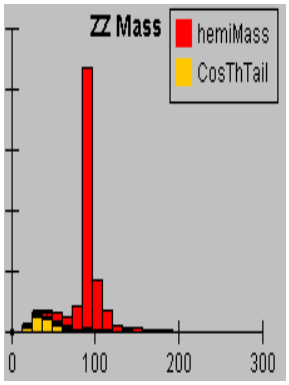

Figure 11: Explanation of bi-modal Z mass distribution.

4.1 Generator level

Table 1 shows the results obtained for the generator level events. In the upper part of the table we see that the qq efficiency cuts select 91% of the qq events while allowing contamination by 36% of the ZZ and only 2% of the tt events. The ZZ efficiency cuts select 84% of the ZZ events and allow 30% of the qq events to pass but only 1% of the tt events. In the case of the tt efficiency cuts 94% of the tt events are selected but only 2% of the qq and 3% of the ZZ events pass that cut.

The lower part of Table 1 shows the results of making cuts aimed at reducing contamination of the qq and ZZ samples. At a one third loss of efficiency the contamination of the qq sample by ZZ events can be decreased by almost a factor of four. For about a one quarter loss of ZZ efficiency its contamination by qq events can be decreased by a factor of greater than seven. Since the tt efficiency cuts were allowing very little contamination, no tt purity cuts were defined.

4.2 Detector level

At the time of this study, the tools for detector simulation were still in development and quite crude. Nevertheless, an attempt was made to set lower bounds on the efficiencies possible using only calorimeter information of the crudest kind. Calorimeter clusters were defined by simply lumping together all contiguous hit cells into a single cluster. Any cluster with less than 0.25

Efficiencies and Contaminations
Generator Level Events

	Event samples qq	ZZ	tt
qq	91%	36%	2%
ZZ	30%	84%	1%
tt	2%	3%	94%

Cuts for efficiency

qq	63%	10%	0%
ZZ	4%	61%	1%

Tighter cuts to decrease contamination

Table 1: Efficiencies and contaminations - generator level.

GeV was excluded. The clusters are assumed to point back to the interaction point which, of course, is not correct for charged particles in a magnetic field. The clusters in the electromagnetic section are treated as gammas (massless) and the hadronic calorimeter clusters are assumed to have the pion mass. No tracking information is used. Furthermore, the cuts used for the generator level were used without any retuning for the detector level. Two detector designs are simulated and referred to as the Large and the Small designs.

The Large design is derived from having a large central tracker to maximize tracking resolution. In the barrel region there is a 40 layer lead/scintillator electromagnetic calorimeter just outside the 200 cm radius tracker with 40 mr square cell segmentation and a depth of 28 radiation lengths. Outside the electromagnetic calorimeter is a 120 layer lead/scintillator hadronic calorimeter with 80 mr square cells. The combined calorimeters are 6.6 interaction lengths deep. There are endcap calorimeters with the same parameters closing down to a radius of 25 cm around the beamline.

The Small design is derived from the desire to get the most accurate vertexing information. This requires a vertex detector close to the interaction point which in turn requires a very large magnetic field to keep background beam radiation out of the vertex detector. The strength of the magnetic field limits the radius of the solenoid which in turn limits the size of the tracker and electromagnetic calorimeter which lie inside the solenoid. The barrel electromagnetic calorimeter is 50 layers of tungsten/silicon with 1.5cm by 1.5cm segmentation and a depth of 29 radiation lengths beginning at a radius of 70cm. The

Efficiencies and Contaminations
Large LCD Detector

Event samples

	qq	ZZ	tt
qq	89%	61%	10%
ZZ	30%	63%	6%
tt	3%	6%	67%

Cuts for efficiency

qq	70%	27%	1%
ZZ	1%	6%	0%

Tighter cuts to decrease contamination

Table 2: Efficiencies and contaminations - Large detector.

Efficiencies and Contaminations
Small LCD Detector

Event samples

	qq	ZZ	tt
qq	89%	58%	10%
ZZ	11%	28%	3%
tt	1%	2%	50%

Cuts for efficiency

qq	70%	25%	1%
ZZ	0%	0%	0%

Tighter cuts to decrease contamination

Table 3: Efficiencies and contaminations - Small detector.

hadronic calormeter lies outside the coil. It is 38 layers of copper/scintillator with 40mr by 40mr cells. The combined calorimeters are 6.1 interaction lengths deep. The endcaps are of similar design beginning at 20cm from the beamline.

The results are shown in Table 2 for the Large and Table 3 for the Small design. The degradation from the ideal (generator) level to the crudest (detector simulation) level is not that severe in some cases. The worst degradation is in the level of contamination of some of the samples. For example, comparing generator level with the Large detector, the qq efficiency dropped only 2% from 91% to 89%; however the contamination of that sample by ZZ events increased from 36% to 61% of the ZZ events. The Large design produces somewhat better results than the Small. The worst results are seen with the cuts to reduce backgrounds in the ZZ sample where all the samples including the ZZ have been essentially eliminated.

Again we emphasize that only crudely defined calorimeter cluster information was used and that the cuts were not retuned from the values choosen for the generator level study. It seems reasonable to expect that the generator level results can be more closely achieved with tuned cuts using more sophisticated calorimeter and tracking information. And again we note that the use of a simple set of cuts instead of something like a neural net does not fully exploit the information available.

5 Conclusions

We have investigated the efficacy of various discriminators for distinguishing hadronic jet origins at a lepton collider with 250 GeV beam energy. Each discriminator has some power to select for or against different kinds of events. The discriminators have been combined with a rough set of cuts to estimate efficiencies and purities possible for event identification. These results should be viewed as lower bounds on what will be possible with additional effort. The indications are that good levels of identification should be possible.

Acknowledgments

This work supported in part by the U.S.Department of Energy under Contract DE-AC03-76SF00515. The author wishes to thank Philip Burrows and David Muller for helpful advice.

References

1. H.-U. Bengtsson and T. Sjöstrand, Computer Physics Commun. bf 46 (1987) 43.
2. S. Catani et al., Phys. Lett. bf B263, 491 (1991).
3. S. Brandt et al., Phys. Lett. **12**, 57 (1964), and E. Farhi, Phys. Rev. Lett. **39**, 1587 (1977).