Measurement of Z boson pair production and a search for the Higgs boson in e+e- collisions at LEP

van Vulpen, I.B.

Publication date
2002

Chapter 4

Standard tools

In this section a few of the tools that are used when studying multi-jet final states are introduced and discussed shortly. The section ends with a description of the selection criteria used to select a pure sample of 4 quark final states from all e^+e^- interactions. This forms the starting point of both the fully hadronic ZZ cross section measurement and the search for the SM Higgs boson.

4.1 Production of multi-jet final states

The quarks that are produced in the e^+e^- interaction can not be observed freely in nature. Due to the colour confinement in QCD they will fragment and form jets of colour neutral particles, allowing quarks and gluons only to be observed through their remnants in the form of (collimated) jets of particles. The translation from jet-characteristics back to partons-characteristics is described in section 4.2. Inversely, the fragmentation from partons to jets, as shown schematically in figure 4.1, is shortly discussed below.

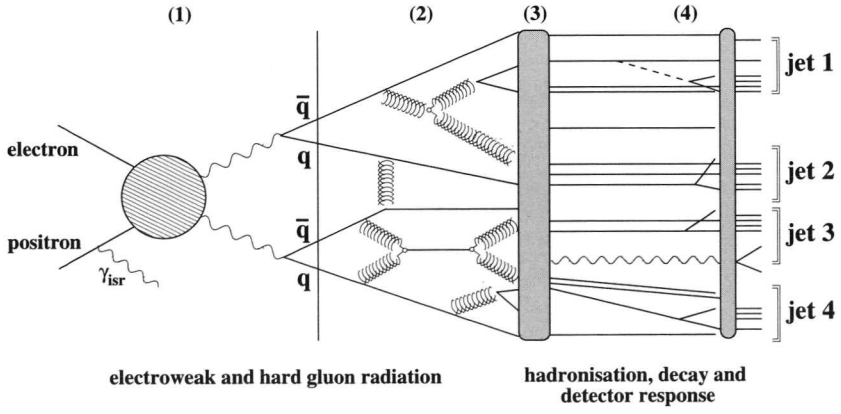

Figure 4.1: *Schematic representation of an* e^+e^- *interaction producing a multi-jet final state.*

In the process leading from the initial e^+e^- system to a set of (collimated) jets of particles, various physics processes and regimes are encountered. To obtain an optimal description of the full pro-

cess, it is divided in parts that can either be computed accurately or have to be modelled. These different regimes are then matched together in a way similar to that shown in figure 4.1. Starting with the incident leptons (on the left) we can identify:

phase 1 & 2) producing a 4 quark final state and hard gluon radiation.

The first two steps in the process are well understood. The electroweak process describing the e^+e^- interaction producing a boson (pair) in the final state (either W^+W^-, ZZ or ZH) and their subsequent decay is well known and the matrix elements associated to specific final states can be computed to high precision. For boson pair production (4-fermion events) the EXCALIBUR [34] programme is used to generate events whereas for the annihilation processes ($e^+e^- \rightarrow Z \rightarrow q\bar{q}$), producing 2-fermion events, the PYTHIA [41] Monte Carlo programme is used.

The second step of the process describes subsequent hard gluon radiation of the initial partons. This region of large momentum transfers is the regime of perturbative QCD. The probabilistic evolution calculations, describing splitting of quarks and gluons, are performed inside the JETSET programme [41]. Hard gluon radiation is the reason that also 2-fermion events produce apparent 4-jet final states. The specific characteristics of the gluon radiation, infra-red and collinear divergences, allow to separate between 4-fermion (spherical) and 2-fermion (more cigar-like) events on a statistical basis. Such a separating variable is constructed in section 4.5.2 and is used both in the pre-selection of multi-jet final states and in the full analysis. This parton shower evolution is cut off when the virtuality of the partons is of the order of about 1 GeV. At that stage there are still only a limited number of partons in the final state.

phase 3 & 4) hadronisation, decay and detector response.

At scales of around 1 GeV QCD perturbation theory breaks down and the picture of QCD in terms of 'free' quarks and gluons no longer holds. This means that we have to use non-perturbative (phenomenological) models to make the transition from a few coloured partons to a jet consisting of many colourless hadrons. This is the step in the (multi-)jet production that can not be computed and there are different models that have been developed to describe the transition. Inside the JETSET programme for example the (Lund) string fragmentation model is used [41]. The hadronisation parameters in this model are tuned on the millions of Z events collected at LEP1 and the combination of parton showers and string fragmentation has been very successful in describing hadronic Z events.

Unstable hadrons formed during the hadronisation phase can decay before reaching the detector. Their decay properties are known from measurements and these decays are handled by the JETSET programme using tables of decays modes and branching fractions. Finally, the dynamics and behaviour of the stable particles inside the different detector components is simulated by the DELSIM [54] package which then provides the final topology as reconstructed in the DELPHI detector.

Experimentalists and theorists start from two different sides of figure 4.1 and, despite a not so well understood region called 'non-perturbative' QCD, the kinematics of the reconstructed jet

is known to represent closely that of the underlying quarks or gluons. Only when testing very precise predictions on for example the exact particle content inside the jet or jet (sub-)structure, will the difficulties in the theoretical prediction manifest themselves. In this analysis only a 4-momentum estimate of the underlying fragmenting parton is required and since the overall energy flow of a high energy event is mainly determined by the perturbative process with only a minor additional smearing caused by the hadronisation step, this is expected to be well described by the Monte Carlo.

4.2 Jet clustering

Primary quarks and gluons produced in the interaction are identified as groups of particles in the detector as described in section 4.1. For an experimentalist the task is to traverse figure 4.1 in reverse order: reconstruct the dynamics of the underlying partonic structure given the detected event. Simplifying the event to a few partons allows to focus on a region where the theoretical predictions are well behaved (QED and perturbative QCD). To obtain the best estimate of the energy and momentum of each parton the goal therefore is to cluster all particles originating from the same parton fragmentation. At the end of the clustering the remaining groups of particles are called jets and their 4-momentum (the direct sum of all particles assigned to the jet) is used as an estimation of the parameters of the primary parton. In this section a general introduction on clustering algorithms is given followed by a short discussion on the DURHAM algorithm which is the algorithm used throughout the rest of this thesis.

Clustering algorithms in general

Most clustering procedures can be characterised by three definitions:

- **A distance** (y_{ij})**:** A measure defining 'how far' two particles i and j are apart. This measure is used to decide which two particles should be considered to be combined first.
- **A maximum distance** (y_{cut})**:** A resolution parameter defining the maximum distance for which two particles can be combined (and therefore the moment the clustering is ended).
- **A recombination scheme:** A prescription of how to construct the 4-momentum of the object that is formed when two particles are combined.

In order to cluster particles into jets most algorithms follow a similar procedure: iterative procedures starting for each event with all (n) measured energy flow objects (particles). For each pair of these final-state particles (i, j) their separation is calculated using the algorithm specific distance definition. In the next step of the procedure the two particles (i,j) with the smallest value of y_{ij} are combined and replaced by a 'pseudo-particle' with 4-momentum $\mathrm{p}^{\mu}_{(ij)}$ using a pre-defined recombination scheme. This clustering is performed only if y_{ij} is smaller than the maximum distance (or resolution) parameter y_{cut}. After clustering particles i and j the event now consists of n-1 clusters and again all distances between particles are calculated. This procedure is repeated until all pairs of particles have $y_{ij} > y_{\text{cut}}$. Clusters of particles that remain at this stage are called jets. For a given value of y_{cut} the event is characterised by the number of reconstructed jets.

The original Jade clustering algorithm

One of the first clustering algorithms is the JADE [55] algorithm, where the distance between two particles is defined as:

$$y_{ij} = \frac{2\,E_i E_j (1 - \cos\theta_{ij})}{E^2_{\text{vis}}} \tag{4.1}$$

Here θ_{ij} is the opening angle between the two particles, making the numerator identical to the invariant mass squared of the two particles in case both particles are massless. The distance y_{ij} is transformed into a dimensionless quantity by normalising to the total visible energy squared in the event. Although the recombination scheme originally suggested was adding of the 4-momenta a variety of schemes has been used in combination with the distance defined in equation 4.1:

E-scheme:	$E_{ij} = E_i + E_j$	and	$\vec{p}_{ij} = \vec{p}_i + \vec{p}_j$
E0-scheme:	$E_{ij} = E_i + E_j$	and	$\vec{p}_{ij} = (\vec{p}_i + \vec{p}_j)\, E_{ij}/\lvert\vec{p}_i + \vec{p}_j\rvert$
P-scheme:	$E_{ij} = \lvert\vec{p}_{ij}\rvert$	and	$\vec{p}_{ij} = \vec{p}_i + \vec{p}_j$

The reason why different recombination schemes were introduced were mostly given by theoretical considerations. The P- and E0-schemes for example violate energy and momentum conservation respectively, but have the advantage that the reconstructed jets are massless, which is similar to the approach in many theoretical calculations. An overview of the different considerations to decide on a different scheme, their performances and characteristics can be found in various review articles (like [56]).

The JADE clustering algorithm has some characteristics that cause problems in certain classes of events since two low-energetic clusters will be combined even when they have a large opening angle. In events containing many of these low-energetic clusters these can form a 'phantom' jet leading to an unnatural assignment of particles to jets. To overcome these problems other algorithms were developed that were less sensitive to soft gluon emission like the DURHAM clustering algorithm.

The DURHAM clustering algorithm

The DURHAM or $k_\perp$-algorithm [57] is the most widely used algorithm in multi-jet final states at LEP2. It is closely related to the JADE algorithm, but does not suffer from the problems discussed above by adopting a slightly different distance definition and the Lorentz-invariant E-scheme as recombination procedure:

$$y_{ij} = \frac{2\,\min(E_i^2, E_j^2)(1 - \cos\theta_{ij})}{E^2_{\text{vis}}} \tag{4.2}$$

$$p_{ij}^{\mu} = p_i^{\mu} + p_j^{\mu} \qquad \text{(E-scheme)} \tag{4.3}$$

With θ_{ij} representing the opening angle between the two particles, the numerator represents the transverse momentum squared of the lowest energetic particle with respect to the most energetic particle. The distance is again normalised to the total visible energy in the event and has the

advantage compared to the original JADE distance that it assigns low-energy particles to their closest high energy jet first and is therefore both experimentally and theoretically better behaved.

Another advantage of this clustering algorithm is the fact that the hadronisation of the parton final states can be shown to have, on average, little influence on the jet-rates. This can be evaluated on generator level by studying the differences between parton and hadron level. Differences vary significantly between clustering procedures and are shown to be smallest for the DURHAM clustering algorithm ([58, 59]). Although there is no single *best* clustering algorithm (every algorithm has specific qualities and problems), in multi-jet events at LEP2 the DURHAM clustering algorithm is generally accepted as one of the best in reconstructing the energy flow of the original partons and is therefore used throughout this thesis.

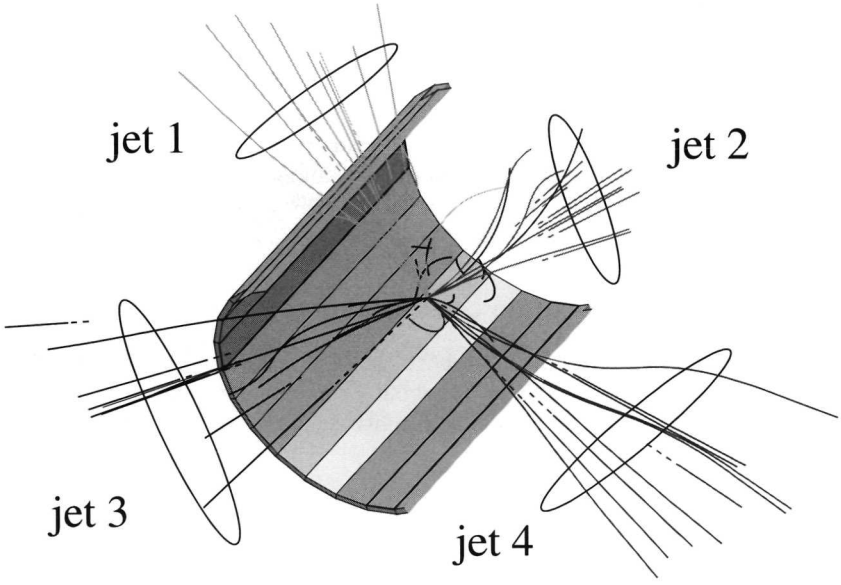

Figure 4.2: *Event display showing reconstructed charged tracks in the DELPHI detector. The DURHAM clustering algorithm has clustered the event into 4 jets.*

The DURHAM algorithm defines the procedure used in this thesis to cluster groups of detected particles into jets whose 4-momentum can be used as an estimate of the original parton 4-momentum. An event in which 4 jets have been reconstructed can be seen in figure 4.2.

4.3 Constrained fit

Estimating the 4-momenta of the partons using the reconstructed jets would be quite accurate (neglecting assignment of particles to the wrong jet) if all particles would be detected and their

energy accurately measured. However, since neutrinos can be present in the final state and particles travelling through inefficient regions of the detector escape detection this is not always the case.

Comparing the reconstructed jet energy to that of the generated parton on Monte Carlo events it is known that, on average, about 15% of the jet energy is lost. The resolution on the jet energies is also about 15% due to both the intrinsic resolution of the calorimetry and the fact that some particles experience significant energy loss through interaction with the detector material. Both effects make accurate jet-energy measurements difficult. The estimation of the parton direction on the other hand is quite precise ($|\theta_{\text{jet}} - \theta_{\text{parton}}| \approx 3°$) since this is a momentum weighted sum of all constituent particles and the resolution on angles and momentum of charged tracks is excellent.

The estimation of the dynamics of the partonic system can be improved by taking into account all known effects and biases. Combining this with the requirement that the reconstructed event should have specific properties (identical to the well-defined characteristics of the initial e^+e^- system) the jet properties can be improved beyond the detector resolution. This will significantly improve the resolution on di-jet invariant masses which is important in studying multi-jet final states. In section 4.3.1 the method (constrained fit) to implement these effects and simultaneously satisfy the constraints will be explained by describing the most general (4C) fit where energy and momentum conservation is demanded. In section 4.3.2 a different (3C) fit will be introduced that is used to provide information for the event selection.

4.3.1 Energy and momentum conservation (4C fit)

Unlike the final state, the initial state of the e^+e^- interaction is known to high precision: perfectly balanced in momentum and energy known to about 1‰ (assuming no initial state radiation (ISR)). Since these characteristics are preserved the reconstructed event has to satisfy 4 constraints (energy and momentum conservation):

$$\sum_{i=1}^{n_{\text{jets}}} p_i^{(\text{fitted})\,\mu} = (0, 0, 0, \sqrt{s}) \tag{4.4}$$

In this fit there are 4 constraints to be satisfied and is therefore known as a 4C-fit. Several methods can be used to take into account the expected energy loss while simultaneously satisfying the different constraints. Here both a rescaling and a more complex constrained fit method are described.

Rescaling: One of the less frequent methods used is the rescaling method characterised by the fact that in the fit the jet directions are fixed and only their energies are allowed to vary. This uses the knowledge that the resolution on the jet-directions is much better than the resolution on the jet energies. This translates into one free parameter per jet and solving for the α_i's in $\sum_{i=1}^{n_{\text{jets}}} \alpha_i\, p_i^{(\text{meas})\,\mu} = (0, 0, 0, \sqrt{s})$. The solutions of this procedure, found by simple matrix inversion, are however not very stable. Especially in planar events (all jets in the same plane) the problems are obvious: in such an event-topology one of the momentum constraints is automatically fulfilled and the solutions become degenerate. Using rescaling to estimate jet energies in events with high invariant di-jet masses, like ZZ or ZH events, introduces large biases making rescaling not the ideal tool to be used in multi-jet events at LEP2.

Constrained fit: To have a stable fit and to be able to take into account the errors on the jet directions the transverse components of the jet momentum must also be allowed to vary. Instead of one parameter this means three independent parameters are needed to parametrise the jet (the invariant mass of the jet is assumed to scale with the energy) in the following way:

$$\vec{\mathbf{p}}_i^{(\text{fitted})} = e^{a_i}\vec{\mathbf{p}}_i^{(\text{meas})} + b_i\,\hat{\mathbf{n}}_1 + c_i\,\hat{\mathbf{n}}_2 \tag{4.5}$$

In this expression $\vec{\mathbf{p}}^{(\text{meas})}$ is the measured momentum and $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ are 2 orthogonal unit vectors in the plane perpendicular to the measured momentum direction $\vec{\mathbf{p}}^{(\text{meas})}$. A schematic view of the different degrees of freedom can be seen in figure 4.3.

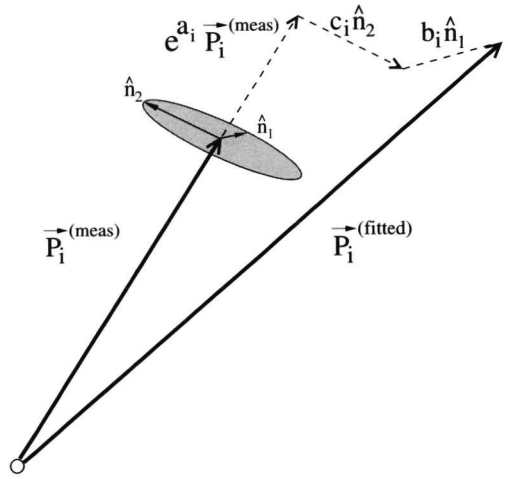

Figure 4.3: *Schematic view of the different components of the jet momentum (a_i, b_i and c_i) that are free to vary in the constrained fit. Both measured and fitted jets are shown.*

This specific parameterisation has been chosen instead of the maybe more obvious set (E,θ,ϕ) because the method used in the constrained fit requires a set of parameters whose errors are described by a Gaussian function. In the fit [60] a Lagrange multiplier technique is used to satisfy the constraints while minimising the following χ^2:

$$\chi^2_{4\text{C}} = \sum_{i=1}^{n_{\text{jets}}} \frac{(a_i - a_0)^2}{\sigma^2_{a_i}} + \frac{b_i^2}{\sigma^2_{b_i}} + \frac{c_i^2}{\sigma^2_{c_i}} \tag{4.6}$$

The parameters to be fitted are a_i, b_i and c_i. The energy loss parameter a_0 ($\sim 15\%$) and the errors on the other parameters (σ_{a_i}, σ_{b_i}, σ_{c_i}) are determined using Monte Carlo events where both the measured and true (parton) momenta are known.

When going to smaller polar angles the resolution on single tracks deteriorates and, due to the geometrical acceptance of the DELPHI detector, particles escape detection. For these reasons the resolutions on the jet-energy and the energy loss of the jet are also polar angle dependent. In

addition, due to the undetected particles at small polar angles, there is a general bias to shift the measured jet directions away from the beam axis. The angular parametrisation of the errors has been tuned using Monte Carlo (2 jets at the Z-pole) events and has been found to be described by the following function and parameters [61]:

$$\begin{aligned} a_0 &= 0.15 + 0.40\cos^4(\theta_i) \\ \sigma_{a_i} &= 0.15 + 0.40\cos^4(\theta_i) \\ \sigma_{b_i} = \sigma_{c_i} &= 1.6 \ + 1.0 \ \cos^4(\theta_i) \end{aligned} \tag{4.7}$$

Using a constrained fit not only results in the best estimate of the parton 4-vectors, but also allows event-by-event errors on these parameters to be used in the analysis. The errors on the jet 4-momenta for example will be used in section 5.2.2 to test the compatibility of a di-jet invariant mass with a specific hypothesis.

4.3.2 Other constrained fits

In addition to the 4C-fit as described by equation (4.4) there are also 3C, 5C and 6C fits (corresponding to the number of constraints they put on the event) that are widely used. In this section the 3C-fit is described since it provides information used in the selection of high-energy multi-jet hadronic events (to reject a large fraction of the $q\bar{q}(\gamma)$ events). The 5C and 6C-fits deal with the extraction of the di-jet mass information contained in the event and can only be discussed when related to a specific physics analysis or jet-pairing. The implementations and characteristics of these fits will be described and discussed in the relevant analysis sections.

3C-fit: Like a 4C-fit, but p_z is left free.
A fit assuming a single photon (γ_{isr}) was emitted along the z-axis. Its momentum is left free. The constraints used in the fit then take the following form:

$$\sum_{i=1}^{n_{\text{jets}}} p_i^{(\text{fitted})\,\mu} = (0, 0, p_z^{\gamma_{\text{isr}}}, \sqrt{s} - |p_z^{\gamma_{\text{isr}}}|) \tag{4.8}$$

Why a 3C fit: To estimate the effective-centre-of-mass energy ($\sqrt{s'}$).
The e^+e^- system can radiate a photon 'before' the interaction (initial state radiation or ISR) thereby reducing the effective centre-of-mass energy in the e^+e^- interaction. These photons are predominantly emitted at very small angles (well inside the beam pipe) and escape detection. Although the energy distribution of these photons is peaked at small energies the distribution has significant tails. In about 12% of fully hadronic 4-fermion events at a centre-of-mass energy of 206.7 GeV for example, the generated ISR energy is more than 10 GeV and for 5% of the events this is even more than 20 GeV.

Although the resolution on the reconstructed $\sqrt{s'}$ is not very good (a few GeV, see also [62]) the $\sqrt{s'}$-distribution can be used to reject events with significant photon radiation. This allows a large fraction of the $q\bar{q}(\gamma)$ events at LEP2 (produced through e^+e^- annihilation) to be discarded already at pre-selection. Since the annihilation cross section is very large at the Z-peak (91.2 GeV/c^2), in about 50% of the interactions a large amount of energy is radiated such that the effective centre-of-mass energy of the e^+e^- system is brought down to the Z-peak. This photon energy emission

pattern results in a characteristic double peak structure in the $\sqrt{s'}$ distribution. Since 4-fermion final states do not have this structure the evaluation of $\sqrt{s'}$ provides a powerful tool to eliminate a large fraction of the $q\bar{q}(\gamma)$ background. The distributions of $\sqrt{s'}$ for different event types can be seen in the right plot of figure 4.10.

4.3.3 Performance of the 4C-fit

In this section the performance of the constrained fit is investigated by comparing, in Monte Carlo events (4-jet events at LEP2), the energy and momenta of the reconstructed jets to that of the generated initial partons. The resolution on the estimation of the angle of the partons is good as can be seen in the left plot of figure 4.4. For jets reconstructed in the central part of the DELPHI detector the performance on the parton energy estimation is shown in the right plot of the same figure. On average the jet-energy is overestimated. This can be explained by the fact that in the constrained fit the initial state radiation is neglected (although $<E_{isr}>$=3 GeV for W^+W^- events at 200 GeV), hence overestimating the energy contained in the event (jets). This also explains the asymmetric distribution and enhanced (ISR) tail at higher (overestimated) jet energies.

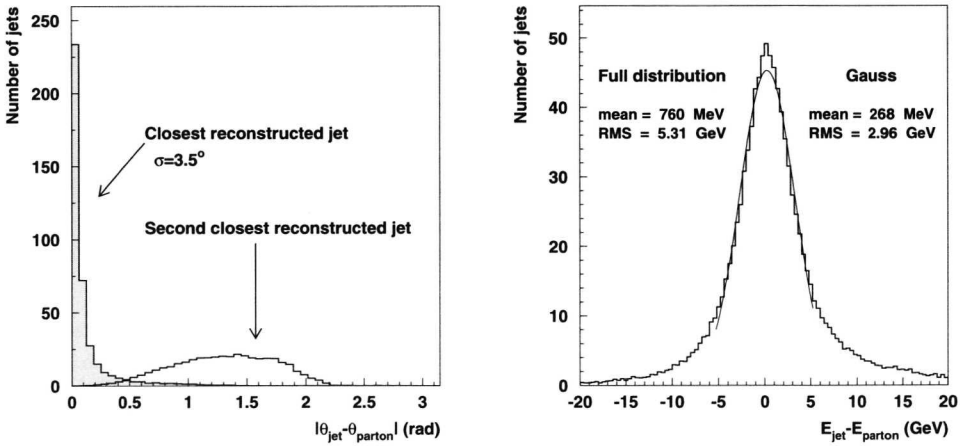

Figure 4.4: *The left plot shows the angle between the original parton and the closest and one but closest reconstructed jet. The right plot shows the energy difference between the reconstructed jet and the energy of the initial parton. These plots were constructed using 4-jet events.*

4.4 b-tagging

B-hadrons are formed in the hadronisation of b-quarks and have quite distinct properties allowing jets of particles to be identified ('tagged') as coming from the fragmentation of a b-quark. The b-tag configuration of different 4-quark final states at LEP2 can be quite specific for a given process. They can therefore be used to select or reject various physics processes (like W^+W^- events where there is hardly ever a b-quark present in the final state). Two complementary methods have

been developed by DELPHI to identify b-quark jets. They have been combined to obtain optimal sensitivity. A detailed description of the method can be found in [63].

Since a b-quark produced in the interaction can not exist freely in nature it will fragment and in the hadronisation phase a B-hadron will be produced. These hadrons decay under the weak interaction via Cabibbo suppressed transitions (at tree level) of a b-quark to either a c- or u-quark under the emission of a W boson. The very small value of the CKM matrix element $|V_{cb}|$ ($\sim$0.04) and the even smaller value of $|V_{ub}|$ ($\sim$0.0035) lead to long life-times of B-hadrons of around 1.6 ps. In combination with the Lorentz boost they acquire at LEP the B-particles will travel several millimetres in the detector before they decay, giving them a characteristic displaced vertex topology as can be seen in figure 4.5. The most powerful technique to identify b-jets is to use the excellent resolution of the vertex detector to search for reconstructed tracks that do not originate from the primary interaction point. This is called the life-time or impact parameter method. However, since hadrons containing a c-quark can also have long life-times this method has limitations.

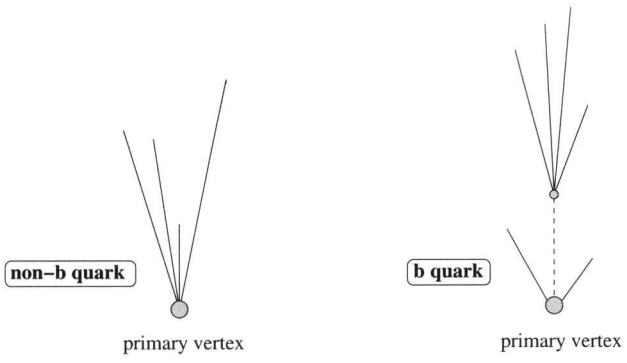

Figure 4.5: *Schematic view of quark fragmentation topologies for non-b (left) and b-quarks (right).*

To improve the b-identification performance beyond the life-time information alone, a second set of variables is introduced, using the specific characteristics of the B-hadron decay itself: the B-hadron is a heavy particle ($\approx$ 5 GeV/c^2), it has a large decay multiplicity and the fraction of the jet energy carried by the charged particles at the secondary vertex is larger for b-quarks than for light (u,d,s,c)-quarks.

Impact parameter or life-time method

Life-time based b-tagging requires testing if (a group of) particles originate from the primary vertex. In that procedure an accurate determination of the primary vertex is of course crucial (see section 3.2.1). Once the primary vertex is reconstructed an *impact parameter* can be defined for each charged particle as being the closest distance between the extrapolated track and the primary

vertex itself. The sign of the impact parameter is defined with respect to the jet direction. It is positive if the extrapolated track crosses the axis of the jet to which it belongs before the primary vertex and negative otherwise.

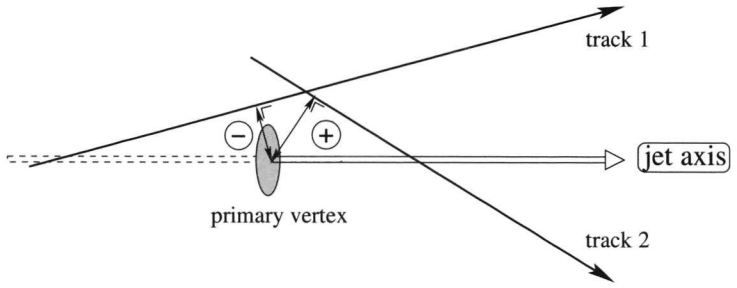

Figure 4.6: *The definition of the (sign of) impact parameter for tracks. Track 1(2) has a negative(positive) impact parameter.*

The resolution on the impact parameter is a measure of the quality of the tracking system. The impact parameter resolution involves a measurement error and a contribution from multiple-scattering. In DELPHI the impact parameter resolutions in the Rϕ-plane and Rz-plane are parametrised as:

$$\sigma^2_{IP\ R\phi} = \left(\frac{\alpha_{\rm ms}}{p\sin^{3/2}\theta}\right)^2 + (\sigma_{0,R\phi})^2 \ , \qquad \sigma^2_{IP\ Rz} = \left(\frac{\alpha'_{\rm ms}}{p\sin^{5/2}\theta}\right)^2 + (\sigma_{0,Rz})^2$$

In this expression p and θ are the momentum and polar angle of the track and the measurement error is given by: $\sigma_{0,R\phi} = 20$ μm. The multiple scattering coefficients in the Rϕ-plane and Rz-plane are given by: $\alpha_{\rm ms}$ ($\alpha'_{\rm ms}$) = 65 (71) μm $\cdot$ GeV/c [48]. The signed distribution of the impact parameter for data and Monte Carlo on the Z-data can be seen in figure 4.7.

Ideally, for particles originating from the primary vertex, the impact parameter distribution is expected to be Gaussian with a width corresponding to the detector resolution. Unfortunately, due to mis-reconstruction and interaction of particles with the detector material, the distribution has large non-Gaussian tails. For particles originating from the primary vertex the distribution is expected to be symmetric. Particles originating from decay of long-lived particles will have (large) positive impact parameters of typically a (few) hundred μm. They will therefore contribute to the positive side of the impact parameter distribution. Before using the impact parameter distribution (as shown in figure 4.7) as a tool to identify particles from B-decay it must be verified that the full detector response is understood and correctly modelled. This is done by tuning the Monte Carlo to describe the negative side of the impact parameter distribution (containing only the combined detector and mis-reconstruction related effects). The tuning procedure [64] includes alignment of detector elements and removal of tracks in the Monte Carlo. It is performed each year using data taken during short dedicated Z runs where large statistics are collected to investigate detector effects and alignment problems that might differ from year to year.

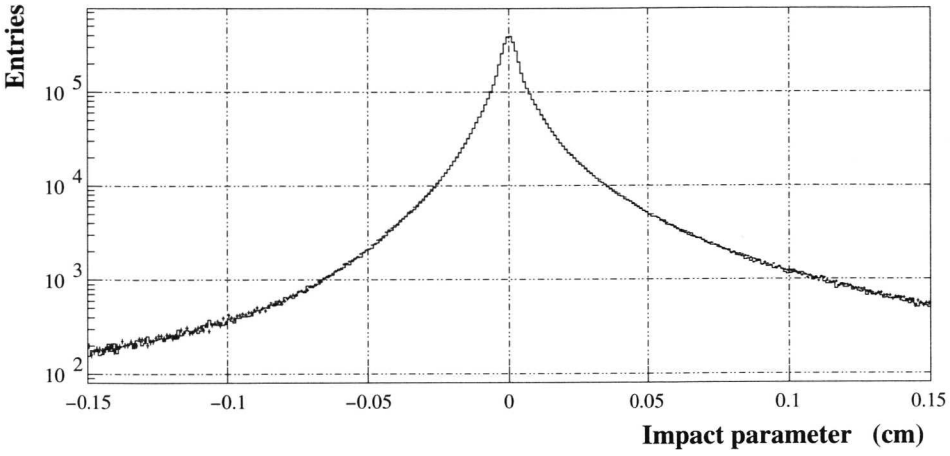

Figure 4.7: *The signed life-time impact parameter distribution after corrections. The points with error bars represent the data, the histogram is simulation.*

Since the precision on the track parameters can differ significantly from track to track it makes more sense to work with the *significance* of a track, defined as the ratio between the impact parameter and its error. Again, the (normalised) distribution of the significance from particles with negative impact parameters, $f(s)$, reflects all combined detector and mis-reconstruction effects and is used as a base to construct the *track probability function* $P(s_0)$, which is the probability for a track from the primary interaction to have a significance with absolute value larger or equal to s_0:

$$P(s_0) = \begin{cases} \int_{s \leq s_0} f(s)ds & \text{if } s_0 \leq 0. \\ \\ P(-s_0) & \text{if } s_0 > 0. \end{cases} \tag{4.9}$$

For tracks originating from the primary vertex, the distribution of $P(s_0)$ is, per definition, a flat distribution between 0 and 1. Tracks from the decay of long-lived particles however have large positive values for s_0 and therefore a small value of $P(s_0)$. This reflects the small probability for tracks from the primary vertex to have such large positive impact parameters. Using the track probability function and the reconstructed significances for any group of N particles the probability can be calculated [65] that they are compatible with originating from the primary vertex:

$$P_N = \Pi \cdot \sum_{j=0}^{N-1} (-\ln \Pi)^j / j! \quad , \text{ where } \quad \Pi = \prod_{i=1}^{N} P(s_i) \tag{4.10}$$

For jets of particles coming from light quark fragmentation the distribution of P_N is expected to be flat, but for jets originating from B-hadron decay this N track probability calculated using formula (4.10) is usually very small due to the significant impact parameters of the decay products of the B-particle. Since a small value of the compatibility with the primary vertex indicates a long life-time this is often referred to as a life-time probability and is the most powerful tool to identify

b-jets. The first global variable that gives information on the b-likeliness will therefore be a life-time one:

- **The jet life-time probability (P_j^+)**
  For each jet in a hadronic event a probability is calculated (using formula (4.10)) that it is compatible with the primary vertex using only tracks in the jet that have positive impact parameters. The distribution of the life-time probability is shown in figure 4.8a for different quark flavours. Due to the non-zero life-time of D-mesons $-\log_{10} P_j^+$ can also be large for c-jets limiting the performance of the life-time tag.

Secondary vertex or B-decay method

To improve the performance beyond the life-time information the characteristics of the B-hadron decay itself can be used. This requires removing the tracks in the jet that do not belong to the B-hadron (see figure 4.5), but are fragments of the b-quark fragmentation in the primary vertex. This is realised by performing a search for a secondary vertex [66] within each jet (the vertex is required to satisfy some quality cuts like the number (or type) of tracks that make up this vertex). In events where a secondary vertex is found additional variables can be defined using exclusively the tracks that make up the secondary vertex.

- **The invariant mass of the secondary vertex particles (M_s)**
  The invariant mass of the particles from the secondary vertex for c-jets is limited by the mass of D-mesons. The probability for a secondary vertex to come from a c-jet decreases sharply above M_s = 1.8 GeV/c^2, while for b-jets the distribution extends up to 5 GeV/c^2 as can be seen in figure 4.8b.

- **The rapidity with respect to the jet direction of tracks from the secondary vertex (R_s^{tr})**
  Although a B-hadron in a b-jet has on average a higher energy than a D-meson from a c-jet the rapidities (defined as $1/2 \ln[(E+p_z)/(E-p_z)]$) of particles from B-decay are on average less than those from D-decay. This can be explained by the larger mass of the B-hadron and the larger multiplicity of its decay [67]. The 'fake' secondary vertices in light quark jets are due to wrongly measured tracks caused by multiple scattering in the detector and interactions with detector material causing the tracks to be shifted to even smaller values. The distribution of this variable can be seen for the three quark classes in figure 4.8c.

- **The fraction of charged energy in the secondary vertex compared to the total jet energy (X_s^{ch})**
  The total energy of charged particles from the secondary vertex in light quark jets is much smaller than in b-quark jets. In case of b-jets the distribution of X_s^{ch} is determined by the fragmentation function $f(b \rightarrow B)$ whereas for a c-quark the distribution is determined by $f(c \rightarrow D)$ which is softer. The distribution of X_s^{ch} can be seen for the three quark types in figure 4.8d.

- **Transverse momentum with respect to the *b*-jet for identified leptons**
  In case the B decays semi-leptonically (Br(B$\rightarrow X l^- \bar{\nu}_l$) $\approx 11\%$) the transverse momentum of the lepton with respect to the B hadron direction [68] can be exploited. Due to the large

mass of the b-quark this transverse momentum is expected to be larger for B-hadrons than for hadrons containing only light quarks. In absence of life-time information this is an excellent separating variable.

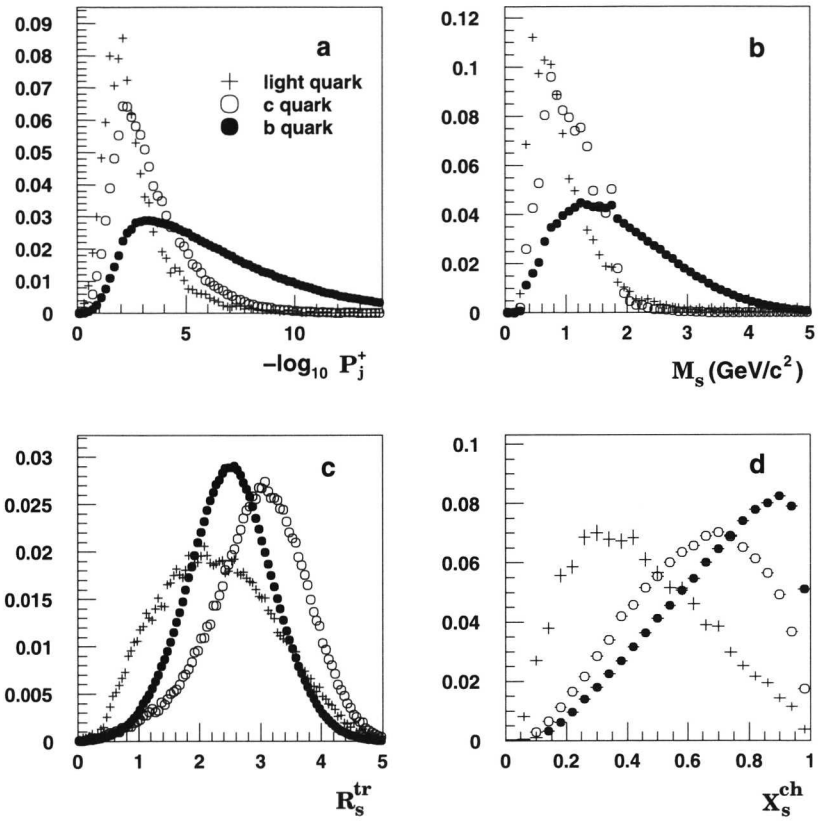

Figure 4.8: *Distribution of the discriminating variables used to construct the combined b-tagging variable. The figures show for different quark types the distributions of: a) the jet life-time probability; b) the invariant mass of particles included in the secondary vertex ; c) the rapidity of particles included in the secondary vertex and d) the fraction of the jet energy carried by the charged particles at the secondary vertex.*

Combining all information and performance

The b-tagging performance is optimised by combining all information on the jet. This is done by defining for each discriminating variable (i) a variable y_i as the ratio of the probability density functions for background and signal respectively: $y_i = f_i^B(x_i)/f_i^S(x_i)$. Since these tagging variables are only weakly correlated, a combined likelihood ratio is defined by simply taking their product: $y = \prod_{i=1}^{n} y_i$.

The backgrounds (corresponding to either c-quarks or u,d,s-quarks) have different behaviour in the variables entering the global b-tagging variable described above. The performance can be improved by taking this into account. This results in the following combined b-jet tagging variable:

$$\rho = n_c \prod \frac{f_i^c(x_i)}{f_i^b(x_i)} + n_q \prod \frac{f_i^q(x_i)}{f_i^b(x_i)} = n_c \prod_i y_i^c + n_q \prod y_i^q \tag{4.11}$$

In this expression n_c and n_q (q=u,d,s) are the normalised numbers of c- and u,d,s-jets with a reconstructed secondary vertex in $c\bar{c}$ and $q\bar{q}$ events respectively (n_c+n_q=1), and $f_i^q(x_i)$, $f_i^c(x_i)$ and $f_i^b(x_i)$ are the probability density functions of the variable x_i in u-,d-,s-,c- and b-quark jets. The jet is tagged as containing a b-quark if $\rho \geq \rho_0$, where the value ρ_0 can be varied to select the desired purity or efficiency of the tagging. The performance of the b-tagging in events with a secondary vertex can be seen in figure 4.9.

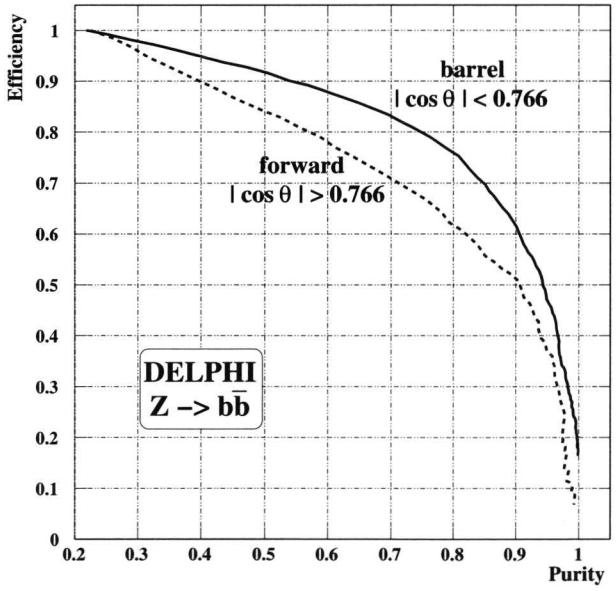

Figure 4.9: *B-tagging (hemisphere) efficiency versus purity.*

$b\bar{b}$ separation

Using more variables (like jet-charge, lepton sign etc.) it is possible to differentiate between jets coming from b- or $\bar{b}$-jets. This could be used to help resolve pairing ambiguities in case there are 2 or more reconstructed b-quarks in the event ($b\bar{b}$ allowed and bb forbidden). In the analysis presented here this possibility is not exploited, but, as will be shown in chapter 5, the structure of the analysis allows inclusion of any new or additional variables in a transparent way.

4.5 Hadronic event selection

In this thesis properties of 4-quark final states are studied. In order to remain as general and flexible as possible there is no optimisation for a specific physics hypothesis[1] during the event selection. Selecting a relatively pure sample of hadronic events requires only a loose pre-selection since its signature is so specific that with only a few cuts (retaining high efficiency for 4-quark final states) backgrounds from other processes are small.

After describing the track selection, the criteria used to select hadronic events are discussed by showing the distributions of variables used, the cuts applied and the data/Monte Carlo agreement before the cuts. In the last section the selection efficiencies for the different SM processes are presented.

4.5.1 Track selection

The observed event is reconstructed using the DELANA programme. A general track fit is performed using charged track segments supplied by all sub-detectors. In this analysis the tracks are required to satisfy the following conditions (default within the SKELANA analysis framework):

- Track momentum ($|p|$) > 100 MeV/c
- Track momentum ($|p|$) $< 1.5 \cdot E_{beam}$
- Relative momentum error (dp/p) $< 1.$

The track should also be roughly compatible with the average primary vertex:

- Impact parameter in $R\phi$ < 4 cm
- Impact parameter in $z \cdot \sin(\theta)$ < 4 cm

There are various classes of tracks that are allowed or rejected. These definitions include the number of hits and the various sub-detectors supplying hits to the track.

A neutral track is defined as an energy deposition in the calorimeter above a specific energy threshold (typically around 350 MeV) that has no charged track pointing to it. To reject beam related background caused by off-momentum electrons there is an additional cut on the polar angle of neutral clusters at 3°.

4.5.2 Event selection

Selection criteria

As a first step the event is required to have a high charged track multiplicity, a large effective centre-of-mass energy ($\sqrt{s'}$: see section 4.3.2) and a large fraction of the available centre-of-mass energy observed in the detector:

[1] In ZH and ZZ analyses a large fraction of the W^+W^- background is often rejected already at pre-selection level by applying an explicit cut on the b-tag variable.

- **cut 1** Number of charged tracks > 17
- **cut 2** $\sqrt{s'}$ $> 0.80\sqrt{s}$
- **cut 3** Total visible energy $> 0.69\sqrt{s}$

The distribution of the number of charged tracks and the effective centre-of-mass energy for various event types can be seen in figure 4.10. These cuts cost almost no efficiency (selection efficiency is 96.0% for fully hadronic 4-fermion events at 200 GeV), but reduce the background significantly (selection efficiency is 16.0 % for non-$q\bar{q}q\bar{q}$ 4-fermion events and 20.2% for $q\bar{q}(\gamma)$ (2-fermion) events at 200 GeV). After these cuts the remaining background is mostly due to $q\bar{q}(\gamma)$ events with hard gluon radiation but also a few non-$q\bar{q}q\bar{q}$ 4-fermion events remain.

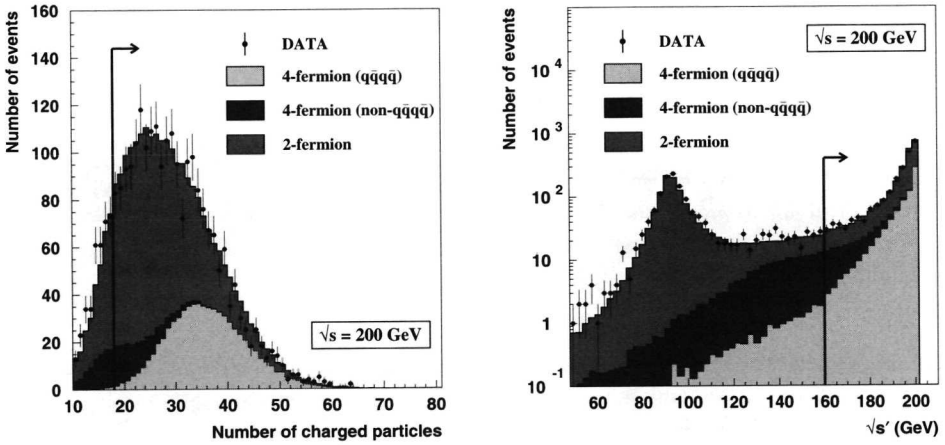

Figure 4.10: *Distributions of variables, at a centre-of-mass energy of 200 GeV, used in the preselection of fully-hadronic 4-fermion events. The plot shows the total charged multiplicity (left plot) and the effective centre-of-mass energy (right plot) when the other two criteria (of the first three) have been satisfied. The line and arrow indicate the region that is selected.*

The events are then clustered into jets using the DURHAM clustering algorithm (y_{cut}=0.001) and events are selected if there are 4 or more reconstructed jets[2]. Due to the kinematics of gluon radiation off quarks ($\sim 1/k_{\text{T}}$) $q\bar{q}(\gamma)$-events have only rarely 4 (or more) well separated reconstructed jets. In addition the knowledge that a quark manifests itself as a (large) number of particles in the detector can be used to reject jets composed of only one (or a few) particle(s) that are often isolated leptons or photons. To reduce the contamination from such events the following jet-quality cuts are introduced:

- **cut 4** Number of reconstructed jets ≥ 4
- **cut 5** Jet quality cuts:
  - Minimum jet multiplicity ≥ 4
  - Minimum jet mass > 1.0 GeV/c^2

[2] Jets with 6 or more reconstructed jets are forced to 5 jets in this analysis.

The distribution of the number of reconstructed jets and the minimum jet multiplicity for various event types can be seen in figure 4.11.

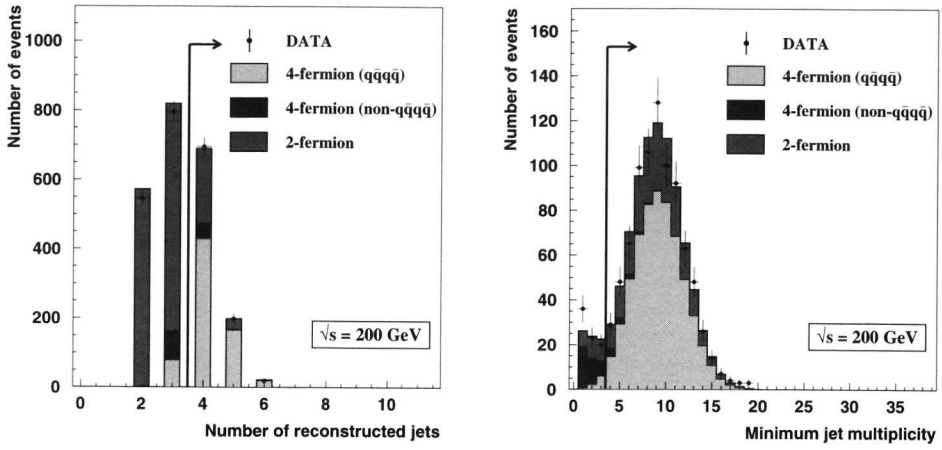

Figure 4.11: *Distributions of variables, at a centre-of-mass energy of 200 GeV, used in the pre-selection of fully hadronic events. The plot shows the number of reconstructed jets (left plot) and the minimum jet multiplicity (right plot) when all other criteria have been satisfied. The line and arrow indicate the region that is selected.*

The final pre-selection requirement uses the fact that the topological structure of the jets (separation between the jets in combination with their energies) in an event is different for 2-fermion ($q\bar{q}(\gamma)$) events with hard gluon radiation and genuine 4-fermion events as described already in section 4.1. The variable used to describe the topology of the event is defined as:

$$D_{pur} = E_1\theta_1\sqrt{E_2\theta_2}/10 \tag{4.12}$$

In this expression E_1 (E_2) is the smallest (one but smallest) energy jet and θ_1 (θ_2) is the smallest (one but smallest) opening angle (in radians) between two jets. Events are required to satisfy the following condition:

- **cut 6** $D_{pur} \geq 7.5$ GeV$^{3/2}$ rad$^{3/2}$

In this section the topology of the event is used to reject some of the clear 2-fermion events, but in section 5.3 it is described in more detail how the full distribution of this variable is used inside the analysis to distinguish between the different event types. In that section also the advantages of using this specific topological variable will be discussed. The distribution of D_{pur} for the various event types at a centre-of-mass energy of 200 GeV can be seen in figure 4.12.

Pre-selection efficiencies

The selection efficiencies for 4-quark final states after all pre-selection cuts is 80% to 90% while most backgrounds are almost completely removed leaving only $q\bar{q}(\gamma)$ events ($\epsilon \approx 2.1\%$) with

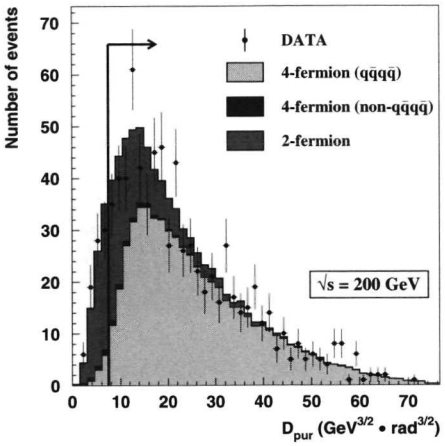

Figure 4.12: *Distributions of the topological (event) variable D_{pur} for the various event types at a centre-of-mass energy of 200 GeV. The line and arrow indicate the region that is selected.*

hard gluon radiation as significant remaining background. The pre-selection efficiencies for the various SM final states are shown in table 4.1.

$\sqrt{s}$	4-fermion $q\bar{q}q\bar{q}$	4-fermion non-$q\bar{q}q\bar{q}$	2-fermion $q\bar{q}(\gamma)$	ZZ $q\bar{q}q\bar{q}$
183 GeV	82.2	1.28	2.12	86.5
189 GeV	82.1	1.29	2.10	88.2
192 GeV	82.5	1.11	1.87	89.3
196 GeV	81.8	1.07	2.00	88.2
200 GeV	81.3	0.91	2.09	87.7
202 GeV	81.7	0.93	2.05	88.2
205 GeV	81.0	0.81	2.10	87.0
207 GeV	80.7	0.93	2.10	87.2

Table 4.1: *Pre-selection efficiencies (in %) for different event types at all LEP2 centre-of-mass energies.*

The pre-selection efficiencies for a two (heavy) boson final state decreases with increasing centre-of-mass energy. This is due to the increasing Lorentz boost the bosons acquire, resulting in an enhanced probability to lose particles in the beam pipe. This in turn results in a slightly enhanced probability for the reconstructed event to be (wrongly) reconstructed as 3 jets instead of 4 or more as can be seen in figure 4.13. This also implies that for a given centre-of-mass energy the pre-selection efficiency for a ZH event depends on the mass of the Higgs boson. For high Higgs masses efficiencies above 90% are obtained. The remaining background from $Z\gamma^*$ events is small

and in signature similar to $q\bar{q}(\gamma)$ events. They will therefore be included in the $q\bar{q}(\gamma)$ numbers in all future plots and tables.

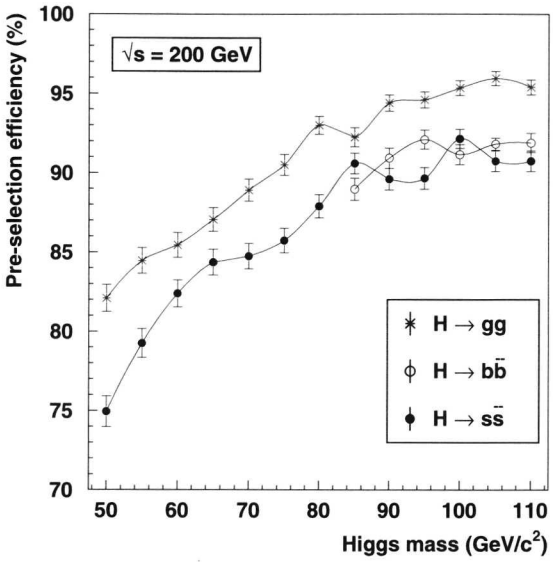

Figure 4.13: *Higgs pre-selection efficiencies as a function of the Higgs mass for different Higgs decay modes. As expected the selection efficiency drops as the Higgs mass decreases since the selection is optimised for heavy boson pair-production.*

Table 4.2 gives a breakdown of the number of expected events from the various physics processes after the pre-selection at the different energies. These numbers are compared with the number of observed events. There is overall good agreement.

$\sqrt{s}$	$\mathcal{L}$	ZZ	W^+W^-	$q\bar{q}(\gamma)$	expected	observed
183 GeV	54.7	6.0	330.5	122.2	458.7	473
189 GeV	158.0	43.6	1015.5	349.9	1409.0	1422
192 GeV	25.9	8.7	169.3	48.9	226.9	239
196 GeV	76.9	29.6	505.6	146.0	681.2	703
200 GeV	84.3	35.2	559.1	160.9	755.2	743
202 GeV	41.1	17.6	271.7	74.4	363.7	368
205 GeV	82.0	34.4	538.3	148.7	721.4	761
207 GeV	142.2	65.2	934.2	250.5	1249.9	1212
total	665.1	240.3	4324.2	1301.5	5866.0	5921

Table 4.2: *The number of expected events after the pre-selection for various physics processes at various centre-of-mass energies. Also the observed number of events at each energy is given.*