Neutral Supersymmetric Higgs Boson Searches at DØ

Stephen Luke Robinson
Imperial College London

A thesis submitted for the degree of
Doctor of Philosophy
of the University of London

July, 2008

Neutral Supersymmetric Higgs Boson Searches at DØ

Stephen Luke Robinson
Imperial College London
July 2008

Abstract

In some Supersymmetric extensions of the Standard Model, including the Minimal Supersymmetric Standard Model (MSSM), the coupling of Higgs bosons to b-quarks is enhanced. This enhancement makes the associated production of the Higgs with *b*-quarks an interesting search channel for the Higgs and Supersymmetry at DØ. The identification of *b*-quarks, both online and offline, is essential to this search effort.

This thesis describes the author's involvement in the development of both types of *b*-tagging and in the application of these techniques to the MSSM Higgs search.

Work was carried out on the Level-3 trigger *b*-tagging algorithms. The impact parameter (IP) *b*-tagger was retuned and the effects of increased instantaneous luminosity on the tagger were studied. An extension of the IP-tagger to use the z-tracking information was developed. A new *b*-tagger using secondary vertices was developed and commissioned. A tool was developed to allow the use of large multi-run samples for trigger studies involving *b*-quarks.

Offline, a neural network (NN) *b*-tagger was trained combining the existing offline lifetime based *b*-tagging tools. The efficiency and fake rate of the NN *b*-tagger were measured in data and MC. This *b*-tagger was internally reviewed and certified by the Collaboration and now provides the official *b*-tagging for all analyses using the Run IIa dataset at DØ.

A search was performed for neutral MSSM Higgs bosons decaying to a $b\bar{b}$ pair and produced in association with one or more *b*-quarks. Limits are set on the cross-section times the branching ratio for such a process. The limits were interpreted in various MSSM scenarios. This analysis uses the NN b-tagger and was the first to use this tool. The analysis also relies on triggers using the Level-3 IP *b*-tagging tool described previously. A likelihood discriminant was used to improve the analysis and a neural network was developed to cross-check this technique. The result of the analysis has been submitted to PRL and is comparable to the result from CDF in the same channel which uses approximately twice the integrated luminosity.

Acknowledgements

I would like to dedicate my work in this thesis to my parents, my gran and grandad and my sister. Without their support and encouragement its hard to imagine how any of this could have been possible.

A huge amount of gratitude must go to Gavin and Per, my supervisors, for all the constructive criticism, enthusiasm and patience they have offered over the years. For learning to cope with my sometimes unusual working practises and for being unfailingly supportive. I would like to thank my undergraduate tutors, Keith, Mandy and Tony, for the wonderful learning environment at St. John's which led me to follow in their footsteps.

I would also like to thank Tim and Miruna who I worked closely with on aspects of my thesis. Both were a pleasure to work with and contributed greatly to my understanding and success. Thanks to Tim, Phil and Daniela, my office-mates while I wrote this thesis, for reading parts of the thesis and responding to my endless questions. I'd also like to thank the other graduate students at Fermilab, especially those who helped me survive without a car and who introduced me to Chicago, Matt, Theo, Ben, Toby, Marilyn, Tamsin and everyone else.

Finally my friends from the 'real' world: Irene, Ashland, Emma, Mark, Miranda, Dan, Joan, John and Naz. Thanks to Isabel and Tahani for wanting to read it.

Contents

List of Figures

List of Tables

Preface

This thesis describes the work performed by the author as a member of the DØ Collaboration primarily between January 2004 and September 2007. The structure of the thesis is as follows:

- **Chapter 1** contains a brief review of the Standard Model (SM) and of the Minimal Supersymmetric extension to the Standard Model (MSSM), with particular emphasis on the Higgs sector. A summary of the experimental status for Higgs searches is also given.
- **Chapter 2** describes the Tevatron accelerator complex and the DØ detector, and introduces some common 'physics objects' used throughout the thesis.
- **Chapter 3** describes development work carried out for the Level-3 trigger, specifically relating to Level-3 b-tagging: the existing impact parameter based b-tagger (IP-tagger) was re-tuned, an extension to the IP-tagger algorithm was developed to include the stereo (z) tracking information (z-IP tagger), and a new secondary vertex based b-tagger (SV-tagger) was commissioned.
- **Chapter 4** covers the certification of the offline neural network (NN) b-tagging tool. This is now the official b-tagging tool used at DØ.
- **Chapter 5** describes two iterations of the search for associated production of MSSM Higgs bosons with one or more b-quarks, where the Higgs decays to a $b\bar{b}$ pair.
- **Chapter 6** gives a summary of the thesis and the outlook for future work.

The key theme of the thesis is the use of b-quarks as a powerful handle on rare events within large QCD backgrounds. Development work was performed on both online and offline b-tagging algorithms and these were ultimately applied to the MSSM Higgs search where the identification of b-quarks is essential for any hope of finding a signal.

Level-3 b-tagging: I was involved in the Level-3 algorithms group from the beginning of my time at DØ. My work has primarily concerned the Level-3 b-tagging algorithms and their dependencies on other tools. The re-tuned IP-tagger was approved by the Level-3 group and used online from summer 2004. Most of the events used in the final analysis were collected using this version of the IP-tagger. An investigation of the effect of changes to the primary vertex and silicon tracker unpacking tools on the IP-tagger performance was a necessary part of the Level-3 group's discussion of these changes, which have subsequently been included in the trigger. The development of the z-IP tagger and SV-tagger offers immediately improved rejection and, in the future, the possibility of a multivariate b-tagger for Level-3 with even better performance. During the development of these tools a mechanism to create a large b-enriched data sample was developed which has subsequently been made available to other members of the Level-3 group. This work is described in detail in Chapter 3. I am responsible alone for all the work in this Chapter except for the development of the SV-tagger which was performed in collaboration with another PhD student.

Offline b-tagging: Before algorithms are offered for use by the Collaboration at large they must be 'certified', for b-tagging tools this means measuring the tagger's efficiency on data and MC for signal and background and producing parametrisations of these efficiencies with errors. The results presented in Chapter 4 refer to the final version of the NN certification. This certified version of the NN provides the official b-tagging for analysis of Run IIa data at DØ. Prior to producing this definitive version of the certification it was necessary to repeat the procedure a number of times due to problems with other offline reconstruction algorithms, corrections to the jet energy scale and to fix problems with the implementation of the NN b-tagger and certification procedure. Additional closure tests and cross-checks were also necessary to confirm the correct implementation of the tool since this certification represents the first collaboration-wide adoption of NN b-tagging. The certification work was carried out with the help of another PhD student. I also produced documentation for the certification code and helped introduce those working on the next generation of certification to the framework.

The MSSM Higgs search: I was involved in two generations of the SUSY Higgs analysis as documented in Chapter 5. The first analysis was presented as a preliminary result at ICHEP 2006. I was primarily concerned with implementing the NN b-tagging in this analysis. I parametrised the data/MC correction factors for b-tagging and optimised the b-tagging working point. This was the first use of NN b-tagging in an analysis at a hadron collider. For the second analysis, which has been submitted for publication in PRL, a new analysis technique was developed by the analysis group. My primary role was again the implementation of b-tagging, the data/MC correction factors and the b-tagging working point optimisation including an investigation into the possibility of using a multi-b-jet event tag. The final submitted result uses a likelihood method as a final event selection. To counter concern that the likelihood might not behave as expected, I developed an alternative multivariate selection technique using a neural network. I also re-optimised the event selection for this NN branch of the analysis.

Chapter 1

The Standard Model, Higgs Mechanism and Supersymmetry

This Chapter briefly covers the theory required to motivate and justify the Higgs search performed here. An introduction is given to the relevant parts of the Standard Model (SM) and Supersymmetry (SUSY). More detail on the SM is given in [1, 2] and more complete reviews of SUSY can be found in [3, 4, 5]. Section 1.1 gives an overview of the SM while Section 1.2 gives a brief description of SUSY. The effect of SUSY on the Higgs sector and the current status of SUSY-Higgs searches are described in Section 1.3.

1.1 The Standard Model

The Standard Model is a relativistic quantum field theory for the fundamental particles and their interactions. It has been extremely successful at making predictions in precise agreement with a wide range of experimental observations [6]. The model consists of fermions, the quarks and leptons, and bosons, the force carriers. Within the SM the known forces occur ‘naturally’ as the result of requiring a set of symmetries, called gauge invariance. However within the gauge invariant constraint it becomes impossible to introduce the required mass terms for the observed particles. An as yet unproven mechanism is needed to complete the Standard Model by giving masses to the fundamental particles. The Higgs mechanism and the associated boson constitute the most widely accepted solution to this problem. Sections 1.1.1 and 1.1.2 will demonstrate how requiring gauge invariance can introduce new fields and how mass terms are excluded. The electroweak combined symmetry group is introduced in Section 1.1.3. A description of the Higgs mechanism is then given with an example of how a Higgs field can generate physically acceptable mass terms for existing particles (Section 1.1.4). The behaviour of the Higgs in the Standard Model and limits on its mass are discussed in Sections 1.1.5 and 1.1.6. Finally some of the problems with this version of the Standard Model are mentioned.

1.1.1 Global Gauge Invariance

The Lagrangian formulation of mechanics has been found to produce the clearest model for the development of relativistic quantum mechanics. The terms of the Lagrangian can be read using the Feynman rules to give an idea of the physical phenomena represented by the field theory. The Lagrangian for a free Dirac particle, a massless four component spinor $\Psi(x)$, is given by the Dirac equation:

$$\mathcal{L}_{Dirac} = \overline{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi \tag{1.1}$$

where γ^μ are the 4×4 Dirac matrices and $\overline{\Psi} = \Psi^\dagger \gamma^0$. Strictly speaking, since Ψ is a field, the quantity $\mathcal{L}$ given here is a Lagrangian density, which can be integrated over a region of space. This Lagrangian density will be loosely referred to as simply the Lagrangian. Since all real world observable quantities are eigen-values or functions of $\overline{\Psi}\Psi$ the observables remain unchanged when an arbitrary constant phase is applied to the field Ψ, i.e.:

$$\Psi \to \Psi' = e^{i\theta}\Psi \tag{1.2}$$

$$\overline{\Psi} \to \overline{\Psi}' = e^{-i\theta}\overline{\Psi} \tag{1.3}$$

The Lagrangian, which classically would describe the equations of motion, should be no exception. The Lagrangian given in Equation 1.1 already satisfies this condition, it is invariant under any global phase given to Ψ. The set of transformations described by all values of θ corresponds to all possible rotations in one dimension. In group theory the set of such rotations corresponds to the group called U(1). A choice of θ is called a gauge and invariance under a global transformation of this kind is called global gauge invariance.

1.1.2 Local Gauge Invariance and Quantum Electrodynamics

The global gauge transformation is a special case of a more general set of transformations where θ is allowed to vary ‘locally’, as a function of position in space-time. The transformation of Ψ becomes:

$$\Psi \rightarrow \Psi' = e^{i\theta(x)}\Psi \tag{1.4}$$

The existing free Dirac Lagrangian is not invariant under this local gauge transformation. An additional term appears from the partial differential of $\theta(x)$:

$$\mathcal{L}_{Dirac} \rightarrow \mathcal{L}'_{Dirac} = \mathcal{L}_{Dirac} - \overline{\Psi}\gamma^{\mu}\partial_{\mu}\theta(x)\Psi \tag{1.5}$$

If local gauge invariance is required the invariance of the Lagrangian can be restored by replacing the partial differential, ∂_{μ}, with a new operator, D_{μ}, called the co-variant derivative. This operator is defined so as to transform in such a way as to cancel out the additional term in the transformed Lagrangian above:

$$D_{\mu}\Psi \rightarrow D'_{\mu}\Psi' = e^{i\theta(x)}D_{\mu}\Psi \tag{1.6}$$

An operator which transforms in this way can be constructed by adding a new real gauge field, $A_{\mu}(x)$, as shown:

$$D_{\mu} \equiv \partial_{\mu} + ieA_{\mu} \tag{1.7}$$

where A_{μ} is required to transform under local gauge transformations as follows:

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} - \frac{1}{e}\partial_{\mu}\theta(x) \tag{1.8}$$

The new Lagrangian formulated with the co-variant derivative is now local gauge invariant, and includes the coupling of the Dirac field to the new spin-1 field A_{μ}.

$$\begin{aligned} \mathcal{L}^{LGI}_{Dirac} &= \overline{\Psi}(i\gamma^{\mu}D_{\mu} - m)\Psi & (1.9)\\ &= \overline{\Psi}(i\gamma^{\mu}\partial_{\mu} - m)\Psi - e\overline{\Psi}\gamma^{\mu}\Psi A_{\mu} & (1.10) \end{aligned}$$

where *LGI* indicates that the Lagrangian has been made locally gauge invariant.

To complete the combined Lagrangian for the Dirac and massless vector fields kinetic and mass terms for the vector field must be added. The appropriate Lagrangian for a vector field is called the Proca Lagrangian. Using the antisymmetric field tensor $F_{\mu\nu}$,

$$F_{\mu\nu} \equiv \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}, \tag{1.11}$$

the following gauge invariant kinetic term can be added:

$$\mathcal{L}_{\gamma}^{KE} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.12}$$

This kinetic term of the Proca Lagrangian can be considered as an alternative formulation of Maxwell's equations. The mass term for such a field is given by:

$$-\frac{1}{2}m_{\gamma}^{2}A_{\mu}A^{\mu} \tag{1.13}$$

This mass term is not local gauge invariant, however if the field is required to be massless the term goes to zero and local gauge invariance is restored for the full Lagrangian. By requiring local gauge invariance a massless vector field which interacts with the Dirac field has been introduced. From the form of the kinetic term in the Proca Lagrangian it can be seen that this free massless field is the photon. The interaction between Dirac particles (electrons, positrons etc.) and electromagnetism (photons, E and B fields) has been introduced 'for free', i.e. without the need to add anything except gauge invariance to the theory by hand. The full Lagrangian for Quantum Electrodynamics is therefore given by:

$$\begin{aligned}\mathcal{L}_{QED} &= \overline{\Psi}(i\gamma^{\mu}D_{\mu}-m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &= \overbrace{\overline{\Psi}(i\gamma^{\mu}\partial_{\mu}-m)\Psi}^{\text{Free Dirac Lagrangian}} - \underbrace{e\overline{\Psi}\gamma^{\mu}\Psi A_{\mu}}_{\text{Interaction}} - \overbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}^{\text{Free Massless Scalar}}.\end{aligned} \tag{1.14}$$

The U(1) symmetry group is the simplest unitary group. By applying local gauge invariance under transformations corresponding to more complex groups, SU(2) and SU(3), the interactions for the weak and strong forces are generated analogously. The number of generators of each group corresponds to the number of bosons required.

1.1.3 Electroweak Unification

The electromagnetic and weak forces can be introduced simultaneously by requiring gauge invariance under a combined symmetry group $SU(2)_L \otimes U(1)_Y$, this is known as electroweak unification [7, 8]. Three fields are generated by the $SU(2)_L$ transformations in weak isospin **T**. These fields, the W_{μ}^{1}, W_{μ}^{2} and W_{μ}^{3} gauge bosons, have a coupling strength g. The subscript, L, indicates that the fields only couple to the left handed component of the fermion fields. The $U(1)_Y$ transformations in hypercharge, Y, produce a field B_{μ} with a coupling g'. The physical bosons are produced by a mixture of these fields:

$$W^{\pm} \equiv (W^1_\mu \mp iW^2_\mu)\ \sqrt{2} \tag{1.15}$$

$$Z_\mu \equiv cos\theta_W W^3_\mu - sin\theta_W B_\mu \tag{1.16}$$

$$A_\mu \equiv cos\theta_W B_\mu - sin\theta_W W^3_\mu \tag{1.17}$$

where θ_W is the weak mixing angle. The EM charge, Q, is given by $Q = T^3 + Y/2$ where T^3 is the eigen-value of the third component of weak isospin and the couplings are related by $g' = g \tan\theta_W$. As with the simpler U(1) gauge invariance described above mass terms for the new gauge boson fields cannot be made locally gauge invariant. In order for the overall Lagrangian to be invariant the masses of the electroweak bosons must be zero. This result is ideal for the photon which we know to be massless (and similarly for the SU(3) group with its massless gluons) however it is inconsistent with the short-range and characteristically 'weak' nature of the weak interaction. Experimentally the $W^{\pm}$ and Z bosons have been observed to have non-zero masses. A new mechanism is required to introduce appropriate gauge invariant mass terms: one proposed method is the Higgs mechanism.

1.1.4 The Higgs Mechanism

The Higgs mechanism was developed to explain the non-zero masses of the W and Z vector bosons whilst preserving gauge invariance [9, 10, 11]. The following example demonstrates how the interaction of the Higgs and gauge fields can lead to a gauge invariant mass term. For simplicity a U(1) invariant Lagrangian is considered here and the example singlet Higgs field introduced will give mass to the photon. Introducing a complex Higgs doublet into the SU(2) invariant Lagrangian will generate mass terms for the $W^{\pm}$ and Z in a similar way.

A complex scalar singlet field ϕ is defined:

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{1.18}$$

$$\phi^* = \frac{1}{\sqrt{2}}(\phi_1 - i\phi_2). \tag{1.19}$$

The U(1) gauge invariant Lagrangian can be written using this complex scalar singlet field:

$$\mathcal{L} = \underbrace{(\partial^\mu + iqA^\mu)\phi^*(\partial_\mu - iqA_\mu)\phi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{Kinetic Term}} \quad - \quad \underbrace{V(\phi)}_{\text{Potential Term}} \tag{1.20}$$

The field ϕ is not free, instead it is under the influence of a potential, $V(\phi)$. The Higgs mechanism relies on a feature of this potential, a non-zero minimum. For illustrative purposes a simple function which satisfies this condition is chosen for $V(\phi)$:

$$V(\phi) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2. \tag{1.21}$$

The potential V(ϕ) is shown in Figure 1.1 for both $\mu^2 > 0$ and $\mu^2 < 0$. If $\mu^2 > 0$ the minimum of the potential is at $\phi = 0$, a scalar field with mass $\sqrt{2}\mu$ is added to the Lagrangian and the gauge fields remain massless. Alternatively if $\mu^2 < 0$ there is an unstable maximum at $\phi = 0$ and a minimum mapping out a circle in the (ϕ_1,ϕ_2) plane:

$$\phi_{1,min}^2 + \phi_{2,min}^2 = v^2 \text{ with } v = \sqrt{\frac{-\mu^2}{\lambda}}, \tag{1.22}$$

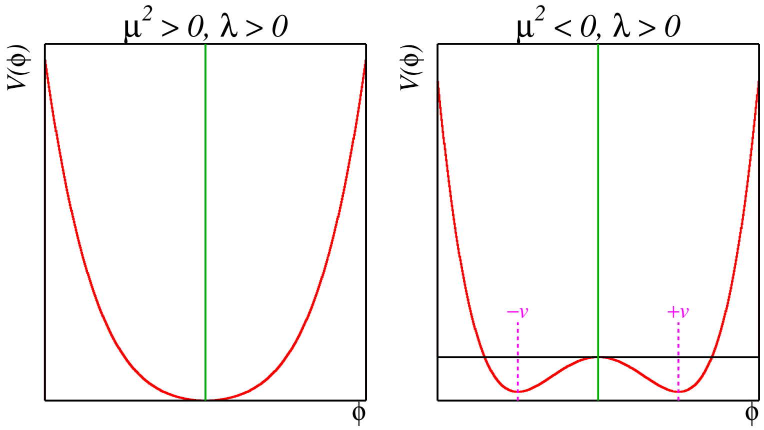

Figure 1.1: The example Higgs potential $V(\phi)$.

where v is the vacuum expectation value. The physical vacuum state can be chosen freely from this set of minima. In choosing a particular value the symmetry is "spontaneously broken". For simplicity the physical vacuum state of the field ϕ is taken to be:

$$\phi_1 = v \tag{1.23}$$

$$\phi_2 = 0 \tag{1.24}$$

Next consider the perturbative expansion of the Lagrangian around this minimum, substituting two new fields defined as:

$$\begin{aligned} \eta &= v - \phi_1 \\ \zeta &= \phi_2. \end{aligned} \tag{1.25}$$

Equation 1.20 can be rewritten in terms of these fields:

$$\begin{aligned} \mathcal{L} \;=\; & \underbrace{\frac{1}{2}(\partial_\mu \eta)^2 \quad - \quad v^2 \lambda \eta^2}_{\text{Massive Scalar Particle}} \quad + \quad \underbrace{\frac{1}{2}(\partial_\mu \zeta)^2}_{\text{Massless `Goldstone' Boson}} \quad - \quad \underbrace{\frac{1}{4} F^{\mu\nu} F_{\mu\nu}}_{\text{K.E. Term}} \\ + \; & \underbrace{\frac{1}{2} e^2 v^2 A_\mu A^\mu}_{A^\mu \text{ Mass Term}} \quad - \quad \underbrace{e v A_\mu \partial^\mu \zeta}_{\text{Interaction Term}} \quad + \quad .... \end{aligned} \tag{1.26}$$

The previously massless gauge field A_μ has acquired a mass term, with the mass of the A boson $m_A = e\nu$. In addition the Lagrangian contains a massive scalar η, with $m_\eta = \sqrt{2\lambda}\nu$, and a massless scalar ζ. The massless scalar is the 'Goldstone' boson [12], a particle expected to be produced with the breaking of any continuous global symmetry. The Goldstone boson is problematic since no massless scalar is observed in nature. The interaction term between A_μ and ζ is also undesirable. However since the physical observables are required to be gauge invariant a gauge transformation can be applied as necessary to remove these problems. The choice of the unitary gauge defined as:

$$\theta = -\tan^{-1}(\phi_2/\phi_1), \tag{1.27}$$

results in the disappearance of terms involving ζ and the full Lagrangian can now be written with $\eta \to h$, the physical Higgs boson:

$$\begin{aligned} \mathcal{L} \;=\; & \frac{1}{2}(\partial_\mu h)^2 \;-\; \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \;-\; v^2 \lambda h^2 \;+\; \frac{1}{2} e^2 v^2 A_\mu A^\mu \\ -\; & \lambda v h^3 \;-\; \frac{1}{4} \lambda h^4 \;+\; \frac{1}{2} e^2 A_\mu^2 h^2 \;+\; v e^2 A_\mu^2 h. \end{aligned}$$

From the initially massless A_μ and complex scalar ϕ two real massive fields have been produced. The A and Higgs fields acquire masses of $m_A = e\nu$ and $m_h = \sqrt{2\lambda}\nu$ respectively. The degree of freedom represented by the complex component of the field ϕ has been absorbed by A_μ to allow it non-zero mass, leaving a real Higgs field h.

1.1.5 The Standard Model Higgs

In the Standard Model the Higgs does not couple to the U(1) gauge field, the photon, as described above. Instead a complex doublet of fields, $\mathbf{\Phi}$, is introduced in the same way and couples to the SU(2) fields, the $W^{\pm}$ and Z bosons. Of the four degrees of freedom in the complex doublet of fields three are transferred to the weak bosons giving them mass and one real Higgs field, h, remains. The process of giving masses to the SU(2) fields within the unified electroweak system is called electroweak symmetry breaking (EWSB). In the Standard Model there is no singlet or triplet of Higgs fields and so no Higgs interactions with the U(1) and SU(3) bosons, no gauge invariant mass terms can be formed and photons and gluons remain massless.

Although the Standard Model Higgs has been introduced as a mechanism for giving the weak bosons mass, the doublet of Higgs fields will also interact with the doublet fields of the fermions. This is called the Yukawa interaction and will be of the form:

$$\mathcal{L}_{Yukawa} = -G_f(\overline{\Psi}_L \mathbf{\Phi} \Psi_R + \overline{\Psi}_R \mathbf{\Phi}^{\dagger} \Psi_L), \tag{1.28}$$

where the Yukawa couplings G_f are free parameters of the theory and $\Psi_{L,R}$ are the left and right handed components of the fermion fields. By choosing the unitary gauge and substituting in the Higgs field with its non-zero expectation value the fermion mass term and a fermion-Higgs interaction term are produced. For example, the Yukawa term for the electron is:

$$\mathcal{L}_{Yukawa} = -m_e(\overline{e}_L e_R + \overline{e}_R e_L) - \frac{m_e}{v}(\overline{e}_L e_R + \overline{e}_R e_L)h. \tag{1.29}$$

The mass is related to the Yukawa coupling by $m_f = G_f v/\sqrt{2}$. Similar terms are produced for the quarks. For all the fermions the strength of the coupling of the fermion-Higgs interaction term is proportional to the fermion mass. For this reason a Higgs below the W^+W^- threshold is expected to decay predominantly to b-quarks with the second largest fraction decaying to τ leptons.

1.1.6 Constraints on the Mass of the Standard Model Higgs Boson

The Higgs mass can be calculated from the v and λ parameters which describe the Higgs field. However λ is a free parameter in the theory and so the Higgs mass is unknown. There are however some theoretical limits as well as direct and indirect experimental limits.

For the Higgs mechanism to function λ must remain positive and finite up to the energy scale where new physics begins. A study [13] has shown that λ can obey this constraint and the Higgs mechanism can continue to function up to the Plank scale if the Higgs mass is between 130 and 190 GeV. In addition unitarity constraints in longitudinal WW scattering processes require that any Higgs in the SM and possible extensions must have a mass below ~ 1 TeV [14].

The Higgs will affect the other Standard Model processes via loop corrections and so measurements such as the W, Z and top masses and widths provide indirect constraints on the Higgs mass. An upper limit is set at $m_h < 190$ at 95% CL using combined measurements from LEP, SLC and the Tevatron [15]. Direct searches at LEP place a lower bound of $m_h > 114.4$ GeV [16].

1.1.7 Problems with the Standard Model

Despite the need for the the Higgs to be light, at least below ~ 1 TeV, loop corrections tend to push the Higgs mass to the highest energy scale at which the Standard Model is valid, perhaps as high as the Plank scale. By fine tuning parameters it is possible to arrive at an acceptable Higgs mass but there is no reason for such a situation to arise naturally. The divergence of the electroweak and Plank scales and the need to fine tune the Higgs mass in the existing theory are known as the 'hierarchy' and 'fine tuning' problems.

The coupling constants of the three fundamental gauge fields 'run' as the energy scale of the interaction varies. As the energy scale increases the three numbers initially appear to tend towards convergence. However the three parameters never converge to a single value. In Grand Unified Theories (GUTs) which aim to unify the three gauge fields, the parameters are required to converge at some high energy 'GUT scale'. To ensure convergence of the gauge interactions at the GUT scale new physics must be introduced between the SM and GUT scales.

1.2 Supersymmetry

Supersymmetry is a proposed symmetry between bosons and fermions. Each particle gains a supersymmetric partner, a 'sparticle', bosons have fermion super-parters and vice-versa. The sparticles have identical quantum numbers to their corresponding particles except for spin. In the observable universe Supersymmetry must be broken since no such super-partners have been observed. This symmetry breaking is assumed to have shifted the sparticles to higher masses.

Supersymmetry solves the two problems with the Standard Model mentioned in the previous section. Loop corrections from super-partners cancel the divergent corrections from the Standard Model particles. The values of the running coupling constants of the gauge fields are also effected by the presence of sparticles above the SUSY scale and all three constants converge at an energy of $\sim 10^{16}$ GeV [3].

1.3 The Higgs in the Minimal Supersymmetric Standard Model

In supersymmetric models additional Higgs doublets are required to maintain U(1) invariance and cancel triangle anomalies [3]. The simplest extension to the Higgs sector, the addition of a second Higgs doublet, is chosen for the Minimal Supersymmetric Standard Model (MSSM). As in the SM each doublet has four degrees of freedom, giving a total of eight. Three degrees of freedom are taken up in EWSB giving masses to the W and Z bosons. The remaining five degrees of freedom produce five Higgs bosons; two neutral CP even, h and H; one neutral and CP odd, A; two charged H^+ and H^-. In the limited parameter space of the MSSM the masses of this set of bosons calculated at tree level can be written in terms of only two parameters, as discussed below. This simplicity is one reason the model is attractive as a benchmark for analyses.

The pair of Higgs doublet fields can be separated into one which couples only to up-type fermions and one which couples only to down-type fermions. The actual Higgs bosons will be a superposition of these fields.

The ratio of the vacuum expectation values of these two fields is equal to $\tan\beta$ where β is the mixing angle between the fields used to construct the charged Higgs states:

$$\tan\beta = \frac{v^u}{v^d} \tag{1.30}$$

where $v^{u,d}$ are the vacuum expectation values of the two up and down-type Higgs fields. The two free parameters needed to completely describe the Higgs masses at tree level are chosen to be $\tan\beta$ and one of the masses, m_A. The CP-even Higgs boson masses are given in terms of the Z mass and the free parameters by equation 1.31.

$$m^2_{h,H} = \frac{1}{2}\left(m^2_A + m^2_Z \mp \sqrt{(m^2_A + m^2_Z)^2 - 4m^2_A m^2_Z \cos^2 2\beta}\right) \tag{1.31}$$

The CP-even masses have further constraints, again at tree level:

$$m_h \leq m_Z|\cos(2\beta)| \tag{1.32}$$

$$m_H \geq \sqrt{m_A^2 + m_Z^2 \sin^2 2\beta} \tag{1.33}$$

Radiative corrections increase m_h^{max}, the upper limit on m_h, from the Z mass to $\sim$ 130 GeV. For large values of $\tan\beta$, $\cos^2 2\beta$ tends to 1 and Eqn. 1.31 gives one of h and H with a mass $\approx m_A$. Three scenarios occur depending on the size of m_A with respect to the upper limit, m_h^{max}:

$m_A < m_h^{max}$ *:* $m_H \approx m_h^{max}$ and $m_h \approx m_A$

$m_A > m_h^{max}$ *:* $m_H \approx m_A$ and $m_h \approx m_h^{max}$

$m_A \approx m_h^{max}$ *:* All three bosons have the same approximate mass.

For larger values of $\tan\beta$ the approximate equalities above tend to become more exact.

Whilst at tree level the Higgs masses and couplings are determined solely by the two parameters already introduced, m_A and $\tan\beta$, higher order corrections depend on further SUSY parameters. Two benchmark scenarios are commonly used to reflect the possible effect of these secondary parameters [17].

m_h^{max} *scenario-* In this scenario the secondary SUSY parameters are chosen so as to give the highest possible value to the lightest of the Higgs masses for a given $\tan\beta$. It is used to give conservative limits on $\tan\beta$.

No Mixing scenario- This scenario assumes a larger scale for the SUSY masses than the m_h^{max} scenario. It is so named because this choice of secondary SUSY parameters suppresses mixing in the t-squark sector.

1.3.1 Production Processes

Figure 1.2 shows the production cross-sections for a SM Higgs [18]. The highest production cross-section for neutral Higgs bosons at DØ is from gluon fusion. For a Higgs produced in this way, decaying to a $b\bar{b}$ pair, the background of $gg, q\bar{q} \rightarrow b\bar{b}$ is overwhelmingly large. Instead searches are performed in channels with associated production of additional heavy flavour jets or weak bosons. In the SM the most common associated production processes are those involving Z or W bosons. Chapter 5 will discuss searches for MSSM Higgs production in association with one

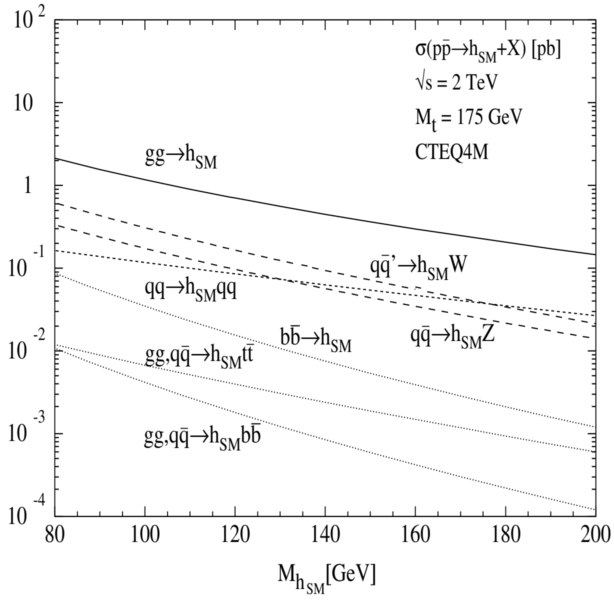

Figure 1.2: The Standard Model cross-sections for Higgs production. Gluon fusion can be seen to dominate over WH/ZH and associated production with b-quarks.

or two b-quarks. These processes are respectively a factor of two and an order of magnitude smaller than the WH and ZH channels in the SM. In the MSSM the coupling of the Higgs bosons to down-type quarks (d,s and b) is enhanced. At tree level the MSSM cross-section for processes where the Higgs is radiated by a b-quark is calculated by scaling the SM cross-section by $\tan^2\beta$. If $\tan\beta$ is large, the production of a Higgs in association with one or more b-quarks can become the dominant production mechanism. Observations at the Tevatron will not allow the CP state of an observed Higgs resonance to be determined. As a result the A, h and H will be indistinguishable and are referred to generically as ϕ or h. The Feynman diagrams for Higgs production in association with one or two b-quarks are shown in Figure 1.3. Following the inclusion of radiative corrections to the tree level diagrams these processes become nearly indistinguishable where three or four jets are observed in the final state. As at least two of the uncharged Higgs bosons are expected at the same mass (within experimental resolution) an overall enhancement of $2\tan^2\beta$ compared to the SM cross-section is expected, as shown in Figure 1.4 [19].

Although values for $\tan\beta$ are unconstrained in theory, a high value is preferred since it would naturally explain the very large mass of the top quark compared to the b-quark [20]. A large value of $\tan\beta$ is also more consistent with the observed density of dark matter [21].

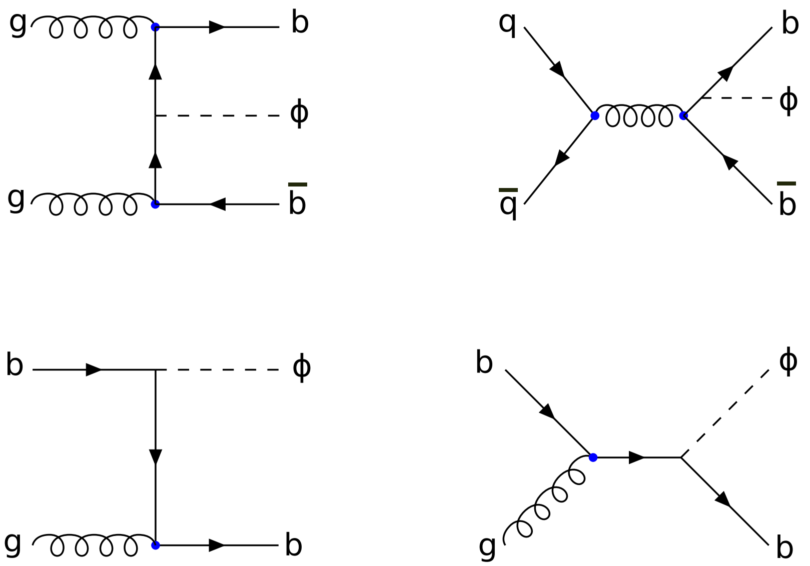

Figure 1.3: Tree level Feynman diagrams showing $gg/q\overline{q} \rightarrow b\overline{b}\phi$ (top) and $gb \rightarrow b\phi$ production (bottom).

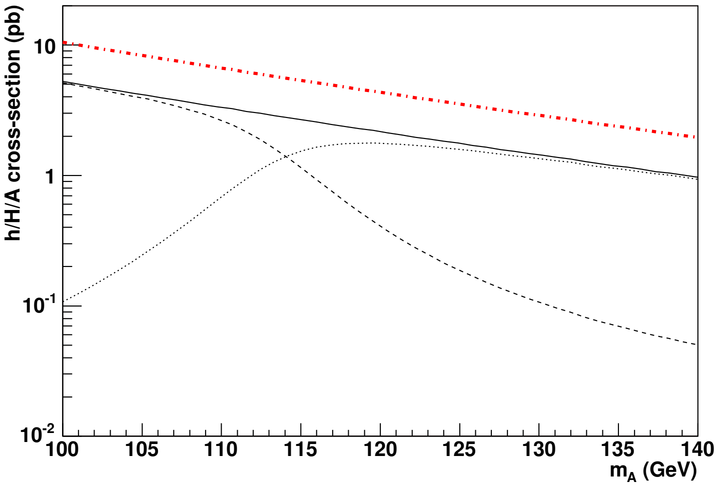

Figure 1.4: Production cross-sections for the neutral MSSM Higgs bosons with a $b\overline{b}$ pair. The calculation is at tree level with $\tan\beta = 30$. The sum of the cross-sections for A (solid), h (dashed) and H (dotted) is shown as the dash-dotted line, equal to twice the cross-section for A.

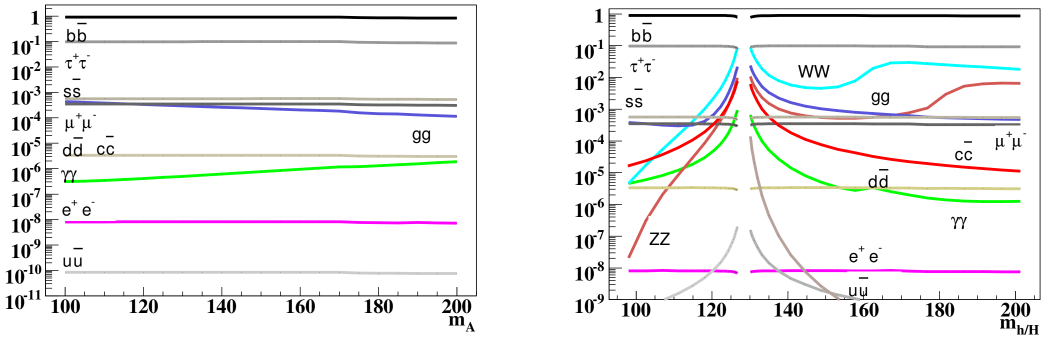

Figure 1.5: Branching fractions of the A (left), h and H (right) in the m_h^{max}scenario of the MSSM for $\tan\beta = 10$.

1.3.2 Higgs Couplings and Decay

For large values of $\tan\beta$ in the MSSM the enhanced Higgs coupling to b-quarks is enough to make $\phi \to b\bar{b}$ the dominant decay channel for all the neutral Higgs bosons and for all values of m_A. The branching fractions for A and h/H are shown in Figure 1.5 for $\tan\beta = 10$ [22]. For all values of m_A, $\sim 90\%$ decay to $b\bar{b}$ and $\sim 10\%$ to $\tau\overline{\tau}$.

1.3.3 Current Status of Analyses

Studies from LEP have excluded Higgs production in the MSSM to 95% CL for $m_{A,h} < 93$ GeV for all values of $\tan\beta$ [23]. Subsequent results from CDF and DØ have further constrained the parameter space excluding the highest values of $\tan\beta$. The preliminary result from DØ in the $b\phi \to bb\bar{b}$ channel [24], which will be discussed in Chapter 5, set a limit on $\tan\beta$ of between 50 and 100 for $100 < m_A < 170$ GeV using $\sim 0.9 fb^{-1}$ and was shown in summer 2006. Subsequent results with $> 1 fb^{-1}$ from the $\phi \to \tau\overline{\tau}$ channel at DØ [25] and CDF [26] have reduced the upper limits to $\tan\beta \sim 40$ for $m_A \approx 140$ and consider higher m_A values (although these are theoretically less likely). A revised version of the $b\phi \to bb\bar{b}$ analysis [27, 22], which uses an increased dataset compared to the preliminary and an improved analysis method, has been submitted to PRL and is discussed in Chapter 5.

Chapter 2

The Tevatron and the DØ Detector

2.1 The Tevatron Accelerator Complex

The Tevatron is a superconducting synchrotron located at Fermilab in the suburbs of Chicago [28]. It collides protons with anti-protons at a centre of mass energy of 1.96TeV on a ring 6.2km in circumference. The collision energy is reached using eight radio-frequency cavities, while 950 superconducting magnets are used to guide and focus the beam. Whilst the Tevatron is the largest and most visible machine at Fermilab it is only the final stage of the larger accelerator complex. It is more efficient to employ a series of accelerators working over specific, limited, energy ranges to reach the desired collision energy. Fig. 2.1 shows a schematic of the multiple stages of the accelerator complex. A full description of the accelerator can be found in [29]. The harmonics of the beams in the Tevatron are such that collisions may be initiated at six equidistant points on the ring. Two of these points are used by collider detectors, one of which is DØ.

2.1.1 Proton Production

Hydrogen gas is ionised with the addition of an electron to form H^- ions. These are accelerated using a Cockcroft-Walton accelerator to 750 keV and then by a Linac to 400 MeV. The ions pass through a fine graphite sheet which strips both electrons leaving the bare protons of the hydrogen nuclei. From here the Booster accelerates the protons to 8 GeV and passes them to the Main Injector where they are accelerated to 120 GeV for anti-proton production or 150 GeV for injection into the Tevatron.

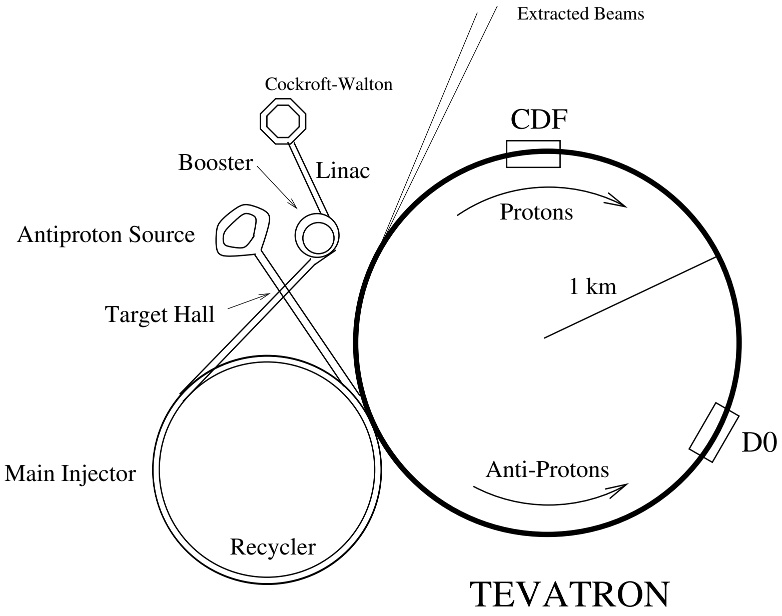

Figure 2.1: Schematic of the Tevatron accelerator complex.

2.1.2 Anti-proton Production

Anti-protons are produced by colliding protons at 120 GeV with a nickel target and sorting the products with a lithium lens and a magnetic field designed to act as a charge-mass spectrometer. For every 10^7 protons hitting the target a single anti-proton at $\sim$8 GeV is produced. Ultimately the number of anti-protons per bunch is around an order of magnitude lower than the number of protons. The Debuncher and Accumulator rings (known as the anti-proton source) are used to collect and cool the anti-protons which are dispersed in energy.

The Recycler ring which occupies the same tunnel as the Main Injector is used for further anti-proton cooling using a beam of electrons. Electron cooling has increased peak luminosities by a factor of two [29]. Using the Recycler as an intermediary stage gives an extra degree of freedom in choosing transfer windows from the anti-proton source to the Tevatron. This allows anti-proton losses at injection to be minimised.

2.1.3 Collisions

The Tevatron simultaneously holds 36 bunches of protons and anti-protons. These bunches are $\sim$ 50 cm long and grouped into three super-bunches separated by 2.6 μs. Within a super-bunch the bunches are separated by 396 ns. Collisions occur at 1.7 MHz. Once a store of proton and anti-proton bunches has been set up it is typically retained for 20 hours.

	Run I	Run IIa	Run IIb
Run period	1993-1995	2002-2006	2006-2009/10
Energy $p\overline{p}$ (GeV)	900	980	980
Proton/Anti-proton Bunches	6	36	36
Protons per bunch	2.3×10^{11}	2.7×10^{11}	2.7×10^{11}
Anti-protons per bunch	5.5×10^{10}	4.2×10^{10}	1.0×10^{11}
Bunch spacing (ns)	3500	396	396
Peak Luminosity ($\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	0.16×10^{32}	1×10^{32}	$\sim 3 \times 10^{32}$
Integrated Luminosity (pb^{-1}/week)	3.2	20	45
Total Integrated Luminosity (fb^{-1})	0.12	1.3	6-8

Table 2.1: Operational parameters for the Tevatron during the three run periods.

Data taking at the Tevatron is divided into three operational periods known as Run I, Run IIa and Run IIb. Run I operated at considerably lower luminosities but is notable (among other things) for the discovery of the top quark [30, 31]. It ended in 1996 to allow for the installation of the upgraded accelerator and detector described here. Run IIa began in April 2002 and ended in March 2006. The integrated luminosity recorded was $\sim$1.3 fb^{-1}. The analysis in this thesis was carried out using the Run IIa dataset. The Tevatron will continue to run until at least autumn 2009 when results from the LHC will begin to become competitive. Currently there is $\sim$ 4 fb^{-1} of data recorded, which gives good prospects for short term improvements to the analysis presented here. By the end of Tevatron running, for Run IIa and IIb combined, we expect to collect 7 - 9 fb^{-1}. Table 2.1 gives the operating parameters for the Tevatron during each of the three run periods.

2.2 The DØ Detector

The detector has a cylindrical structure common among collider based high energy physics experiments. The symmetry of the detector reflects the symmetry of the beam energies. The detector can roughly be divided into three main sub-detectors: closest to the beam pipe are tracking detectors, surrounded by the calorimeter for measuring the energy of electrons, photons and jets, and outermost is the muon system. The original strengths of DØ were its calorimetry and excellent muon coverage. In Run II the addition of a solenoid and silicon tracker has greatly improved tracking resolution and particle identification. For Run IIb an additional tracking

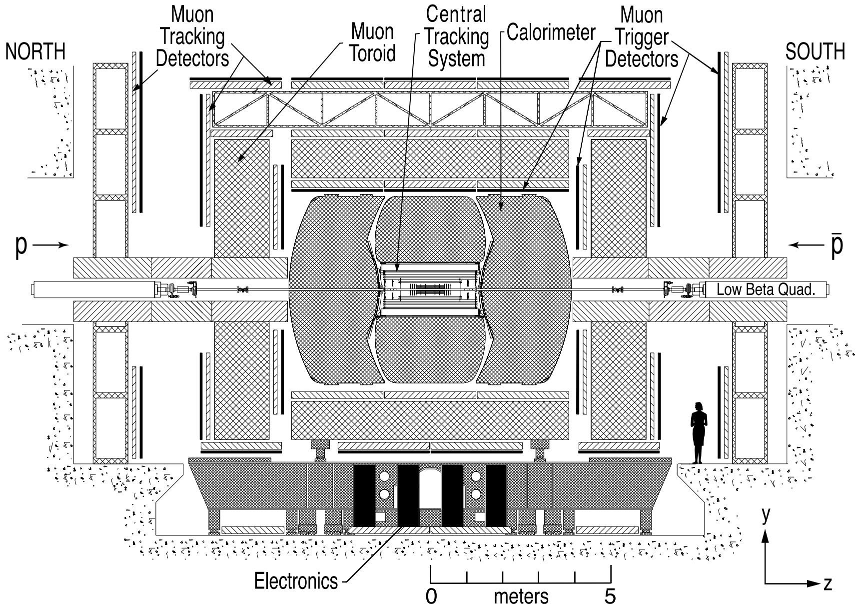

Figure 2.2: The DØ Run IIa detector, cross-section in the y-z plane.

layer has been implemented, along with improved trigger readout electronics. Figure 2.2 shows a cross-section of the detector showing the three sub-detectors. A full description of the detector can be found in [32].

2.2.1 The Coordinate System

The conventional DØ coordinate system has its origin at the nominal interaction point, the centre of the detector. The z-axis points along the direction of the proton beam, the y-axis points up and the x-axis outwards along the radius of the Tevatron. The polar coordinates follow from this with r confined to the x-y plane, perpendicular to the beam axis, and ϕ the azimuthal angle from the x-axis in the x-y plane. The polar angle, θ, from the proton beam direction is usually replaced with the pseudo-rapidity, η. In the limit $E \gg mc^2$ pseudo-rapidity is an approximation to the rapidity which is a Lorentz invariant. Pseudo-rapidity is defined as:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{2.1}$$

The separation of objects in the $\eta - \phi$ plane is usually given in terms of a cone size, ΔR, where:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \tag{2.2}$$

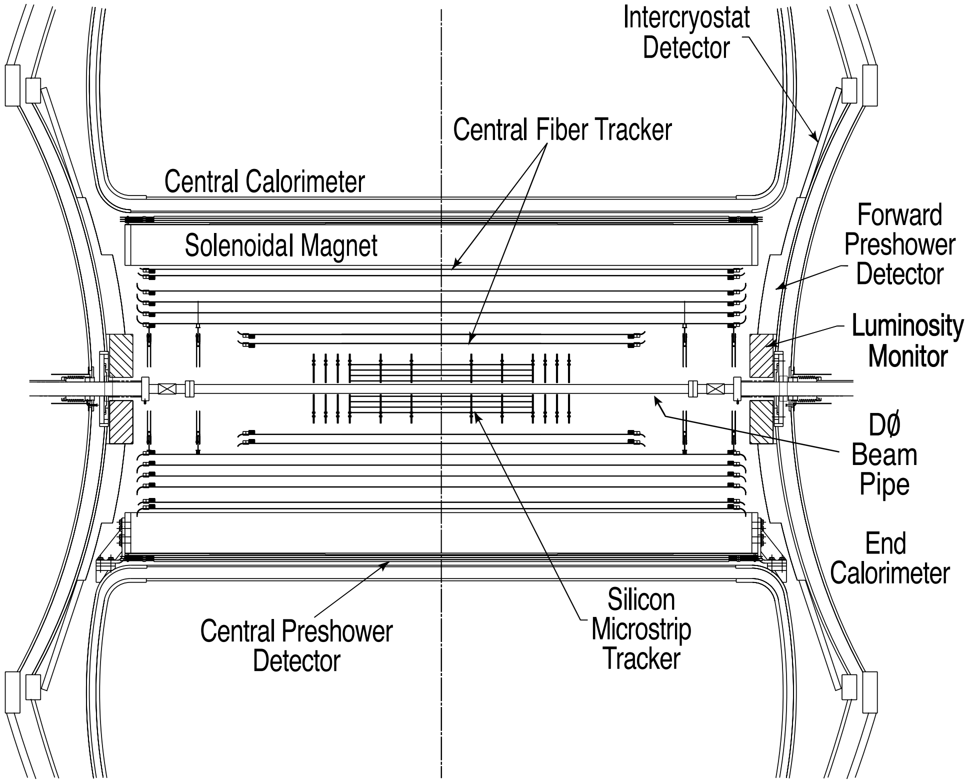

Figure 2.3: cross-section of the tracking system.

2.2.2 The Tracking System

The tracking system is composed of a precision silicon microstrip tracker surrounded by a fibre tracker both contained within a superconducting solenoid producing a 2T field. The system provides tracking out to high pseudo-rapidity ($|\eta| < 3$), primary vertex resolution of 35 μm and a momentum measurement on charged tracks with a resolution of $\delta p_T/p_T = 0.02 \pm 0.0015 \times p_T\%$ where p_T is measured in GeV. For tracks suitable for b-tagging (central η, $p_T > 10$ GeV) the position resolution is 15 μm. The arrangement of the tracker components can be seen in Fig 2.3.

2.2.2.1 Silicon Microstrip Tracker

The Silicon Microstrip Tracker (SMT) is composed of three types of sub-detector. The silicon wafers making up these sub-detectors are held within a Beryllium support structure. The arrangement of wafers can be seen in Fig. 2.4.

Barrels In order to cover the large lateral spread of the interaction point ($\sigma_z \sim 25$ cm) there are three 12 cm barrels either side of the coordinate origin along the z-axis. Each barrel is composed of four double sided layers of rectangular silicon modules,

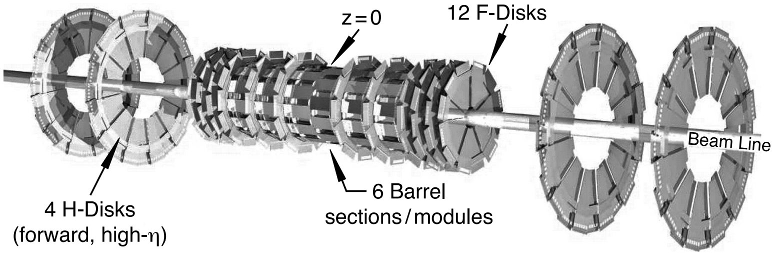

Figure 2.4: The silicon microstrip tracker.

called ladders. The inner layer radius is 2.7 cm and the outer 10.5 cm. In all there are 432 ladders, twelve on the two inner layers and 24 on the two outer layers of each barrel.

F-disks These twelve disks are composed of twelve double sided trapezoidal wedge shaped modules. One disk is attached to the end of each individual barrel. The remaining disks are positioned in two groups of three a short distance from the ends of the complete set of barrels.

H-disks At ∼1 m from the nominal interaction point these offer SMT tracking up to $|\eta| < 3.0$. The H disks are made up of twelve wedges composed of two single sided half-wedges.

The signal to noise ratios for SMT hits in the different detector types varies between 12:1 and 18:1. At the end of Run IIa radiation damage to the SMT had resulted in a loss of $\sim 10\%$ of the detector channels [32].

Since the end of Run IIa the silicon detector has been upgraded with the addition of an extra barrel layer (layer-0) inside the existing inner layer. This layer will compensate for the effect of ageing on the SMT and will in fact improve the b-tagging resolution at low p_T by a factor of two [33].

2.2.2.2 Central Fibre Tracker

The Central Fibre Tracker (CFT) extends out to $|\eta| = 1.6$, providing tracking and allowing a momentum measurement. The tracker consists of eight concentric cylindrical super-layers of scintillating fibres at radii between 20.04 cm $< r < 52.15$ cm. Each super-layer contains two doublet layers of fibres known as the axial and stereo layers. Doublet layers are formed from two sets of 128 fibres overlapping by half a fibre diameter. Axial layers are aligned with the z axis giving $r-\phi$ information.

The stereo layers are alternately at an angle of ±2° to the z-axis and therefore can be used to provide z information. The fibres themselves are polystyrene with a thin double cladding (50 μm on fibres 835 μm in diameter). The core material is doped with 1% p-terphenyl to increase the yield of scintillation light and 1500ppm of 3-hydroxyflavone which shifts the light to a wavelength which is more efficiently transmitted.

The scintillating fibres feed into undoped polystyrene waveguides which take the light to visible light photon counters (VLPCs) located in a cryostat beneath the detector. The VLPCs act as solid state photomultipliers converting the photons to an electrical signal with a fast response and quantum efficiency greater than 75%. For each photon converted the VLPCs output between 22000 and 65000 electrons. At the present time of order 99% of the VLPCs remain functional [32]. Hits in the axial layers provide a fast Level-1 track trigger.

2.2.2.3 Solenoid

The limited space between the calorimeter and tracking is occupied by a solenoid, 2.73 m in length with 1.42 m external diameter. Designed to optimise momentum resolution the solenoidal field is uniform at 2T to within 0.5% within the tracking volume. The material of the magnet and its cryostat are 1.1 radiation lengths (X_0) thick. To allow studies of the systematics the polarity of the solenoidal field is periodically flipped.

2.2.2.4 Preshowers

The preshower detectors (PS) aide in matching calorimeter clusters to tracks and in identifying electrons. A mosaic of triangular polystyrene scintillator tiles is attached to the inner face of the end calorimeters and placed between the central calorimeter and solenoid. The central preshower (CPS) covers $|\eta| < 1.3$ and the forward preshower (FPS) $1.5 < |\eta| < 2.5$.

2.2.3 The Calorimeter

The DØ calorimeter is divided between three separate cryostats, called the central (CC) and north/south end cap (EC) calorimeters. The detectors are further divided into electromagnetic (EM), fine and coarse hadronic (FH/CH) regions, arranged as seen in Fig. 2.5. These regions are composed of multiple layers which are in turn made up of cells. The detector is a sampling calorimeter with each cell consisting of a metal absorber which initiates showering, liquid argon ionising medium and copper charge collection pads. The secondary particles produced during showering in the absorber ionise the liquid argon and under a potential the charge drifts to the pads where it is read out.

The EM region is closest to the beam axis and is designed to contain most of the showering from photons and electrons. The absorber is made of depleted uranium, which acts as a compensator responding almost equally to electromagnetic and hadronic showers. The detector is made up of four layers, the inner three having cells 0.1×0.1 in $\eta - \phi$ and the fourth, where the maximum energy is deposited, having 0.05×0.05 cells. The material of the EM layers presents 20 X_0 in the CC and 21.4 X_0 in the EC.

The FH absorbs most of the energy deposited by hadrons with the CH catching any leakage such that nothing but minimum ionizing particles should in general escape the calorimeter. In total the thickness of the electromagnetic and hadronic regions is 8 - 10 interaction lengths. The hadronic cells have an area 0.1×0.1 in $\eta - \phi$ except for $\eta > 3.2$ where they are $\sim 0.2 \times 0.2$

2.2.3.1 Performance and Status

Since the calorimeter is compensating the EM and hadronic response is almost equal. The ratio (EM/hadronic) varies from 1.11 at 10 GeV to 1.04 at 150 GeV, as measured on electrons and pions in the test beam. However the resolution does depend on particle type and was measured from the test beam [34] as:

$$\sigma_E/E = 0.15/\sqrt{E} + 0.003 \quad \text{for electrons} \tag{2.3}$$

$$\sigma_E/E = 0.45/\sqrt{E} + 0.04 \quad \text{for pions} \tag{2.4}$$

where E is measured in GeV.

Damage to the calorimeter has been limited, indeed only the readout was replaced following Run I. We currently have 99.9% of channels operational.

2.2.3.2 The Inter-Cryostat Region

Scintillating tiles are attached to the inside face of the EC to help recover resolution in the space between the cryostats. The sixteen tiles are divided into twelve readout sectors each covering $\sim 0.1 \times 0.1$ in $\eta - \phi$. Whilst the resolution in this region (known as the ICR) is improved using this detector, it is unavoidably worse than in the main calorimeter volumes.

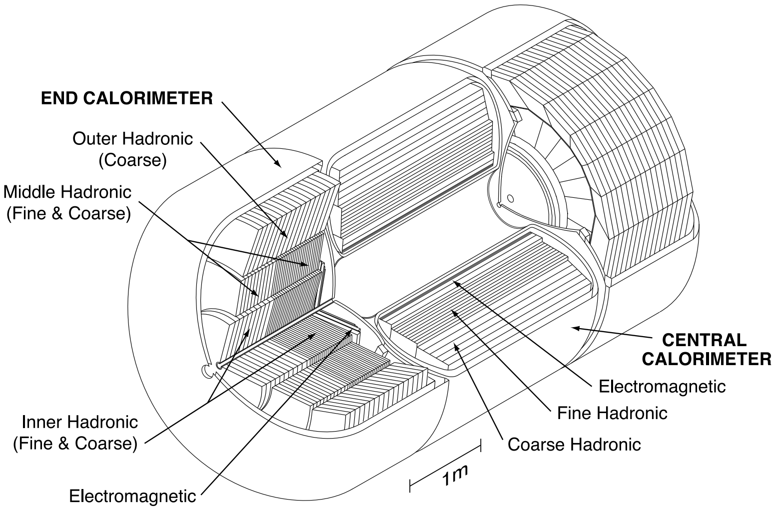

Figure 2.5: The calorimeter.

2.2.4 The Muon System

Outside the calorimeter, the final layer of the detector is the muon system. To detect muons two types of drift tubes are used. Scintillator detectors are also used to veto cosmic muons and to associate muon hits with the correct bunch crossings. The scintillator is also necessary to allow triggering since the drift tube read-out is too slow. The detector is split in two with the central muon system using proportional drift tubes (PDTs) out to $|\eta|< 1$ and the forward muon system out to $|\eta|\sim 2$ using mini drift tubes (MDTs). Each sub-detector consists of three layers of drift tubes, layer A within a toroidal magnetic field, and B/C layers outside it. The magnetic field bends trajectories in the r-z plane allowing a momentum measurement. Layer A has four decks of drift tubes, while the B and C layers have three. The PDTs have a granularity of 10.1 cm and a maximum drift time of 500 ns, while the MDTs are much finer (1 cm cells) and faster (60ns max drift time). Three layers of scintillator are present in the forward system and two are used in the central system.

For low p_T muons the momentum measurement is dominated by the tracking system. The momentum resolution of the muon system is $\Delta p/p = 40\%$ for a momentum of 50GeV. Above 100GeV including the muon system begins to improve the p_T measurement. The position resolution of both types of drift tube is ~ 1 mm. Overall the muon system is over 99% operational, with $< 1\%$ more of the older central PDTs being damaged.

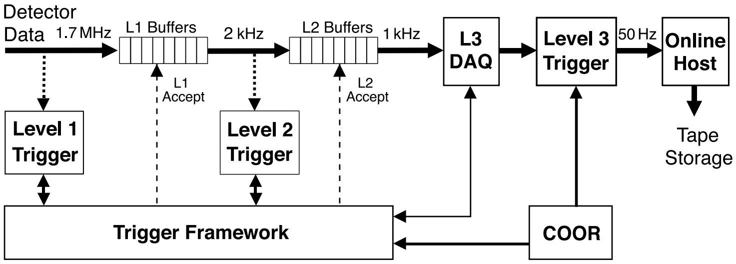

Figure 2.6: The trigger framework.

2.3 The Trigger and Data Taking System

The rate of collisions is far too high to make it feasible to write out and reconstruct all events, furthermore the majority of events are of little interest. A tiered trigger system allows us to perform increasingly complex reconstruction while avoiding pile-up of uninteresting events. Elastic scattering is rejected using the luminosity monitor (LM). The further trigger levels are known as Level-1, -2 and -3, in order of increasing complexity and increasing time available per event (see table 2.2 and Fig. 2.6).

2.3.1 Luminosity Monitor

The LM detects inelastic $p\overline{p}$ collisions using two sets of plastic scintillators at $z = \pm 140$ cm in front of the end calorimeters and between the forward preshowers and beam pipe. The LM covers the region $2.7 < |\eta| < 4.4$. Since the cross-section for all inelastic collisions is well known we can use the rate of LM triggers to calculate the total luminosity (L) where:

$$L = f\overline{N}_{LM}/\sigma_{p\overline{p}} \tag{2.5}$$

where f is the frequency of bunch crossings, $\sigma_{p\overline{p}}$ is the cross-section and $\overline{N}_{LM}$ is the number of LM triggers scaled up to account for the acceptance and efficiency of the detector.

2.3.2 Level-1

The Level-1 trigger decision is taken by hardware; field programmable gate arrays (FPGAs) take partial read out from the PS, calorimeter, CFT and muon detectors. Calorimeter information is processed in 0.2×0.2 $\eta - \phi$ windows called trigger towers. Calorimeter Level-1 trigger terms require either a number of towers over a transverse energy (E_T) threshold or sum all towers and make a cut on the total E_T. The

	Rate	Latency
Collisions	1.7 MHz	n/a
Level-1	1.6 kHz	3.6 μs
Level-2	800 Hz	$\sim$ 100 μs
Level-3	50 Hz	$\sim$ 150 ms

Table 2.2: Output rates for the various levels of the trigger system and the processing time available at each.

CFT triggers look at axial fibres in 4.5° sectors comparing the hits with predefined candidate patterns. The tracking can distinguish between four p_T thresholds and match tracks to preshower hits. The muon triggers consider the candidates from the track trigger and looks for consistent hits in the muon system.

If one of the 128 Level-1 trigger terms (‘bits’) is fired the event is passed to one of 16 buffers to wait for a Level-2 decision.

2.3.3 Level-2

The Level-2 trigger is based on FPGAs like those at Level-1, combined with microprocessors. At Level-2 the simple Level-1 bits are combined with the SMT information to form physics objects (electrons, muons, tracks and jets). Each sub-detector is handled separately and then a further processor, Level-2 global, combines the physics objects to look at the complete picture.

There are one or more Level-2 trigger terms for every Level-1 bit. For each event we run all the Level-2 algorithms corresponding to the Level-1 bits which have fired.

2.3.4 Level-3

The complete output of the detector is read out if Level-2 fires and is available for the Level-3 trigger. Within the constraints of the output rate of Level-2 it performs a partial reconstruction of the event. The algorithms used at Level-3 are similar to the offline tools, but optimised for faster operation. Cuts are kept loose to retain efficiency while keeping to the nominal 50Hz rate to tape. The software consists of tools to process the data and filters to determine which events pass the trigger. For each Level-2 trigger bit fired a filter script is called. The filter script is made up of one or more Level-3 filters which define conditions for passing the event. The filters may act on individual physics objects, on combinations of objects or on combinations of other filter results. If the physics objects required by the filter have not been produced the relevant physics tools are run. Once run the tool's results are cached and made available to other filters. A set of unpacking tools process the raw data readout into more manageable and usable data. The physics tools then make objects or combine objects as requested by the filter scripts. A farm of ordinary Linux PCs is used to run the reconstruction tools and filters. Each farm node processes a single event at a time, independently of the rest. Events which pass the filters are passed to the 'Online Host' to be written to tape.

2.4 Reconstruction Software

A full reconstruction of each triggered event is performed on an offline CPU farm at Fermilab. In many cases the algorithms are similar to those available at Level-3, but here we can afford to take longer where improvements can be made. The complete reconstruction takes around $25s/GHz$ per event at typical luminosities ($\sim 1 \times 10^{32}$ $\text{cm}^{-2}\text{s}^{-1}$). Some of the basic object definitions are described in the following section.

2.4.1 Physics Objects

Primary Vertices Primary vertices are found using an adaptive iterative vertex finding algorithm [35, 36].

Jets Calorimeter jets are formed using a 0.5 cone in ΔR [37]. The jet energy is corrected using the jet energy scale (JES) [38] to recover energy deposited outside the cone and energy not recorded in the calorimeter and to remove other energy sources.

Jet Flavour In our Monte Carlo (MC) samples a jet is called a b-jet if it has a b-quark within $\Delta R < 0.5$. A c-jet is a jet with a c-quark but no b-quark within $\Delta R < 0.5$. The remaining jets, without a c or b-quark within the ΔR cone, are called '$udsg$' or light jets.

Track Jets Track jets are formed from a seed track with $p_T > 1$ GeV and at least two SMT hits. Tracks within $\Delta R < 0.5$ are attached to the seed if they have $p_T >$ 0.5 GeV and at least one SMT hit. Tracks in a track jet must have a small distance of closest approach to the PV (DCA) to ensure they originate from the correct PV. The DCA requirements are $|DCA_{xy}| < 0.2$ cm and $|DCA_z| < 0.4$ cm. A track jet must be composed of at least two tracks [39].

Good Jets We call jets ‘good’ if they are above 15 GeV and within $|\eta| < 2.5$. This is a minimum requirement for jets we expect to be taggable.

Taggability Taggable jets are those good calorimeter jets which are attached to a track jet within $\Delta R < 0.5$. The taggability of a sample is the ratio of taggable jets to good jets [40].

Muons The definition of muon objects used in this thesis requires at least two hits in both drift tube layers, scintillator hits in both the A, and B or C layers, $p_T > 4$ GeV and a track match with $\chi^2 < 100$ [41].

2.4.2 *b*-tagging Algorithms

A jet originating from the hadronisation of a *b*-quark can be identified using the quark’s long lifetime, or for the $\sim 10\%$ which decay semi-leptonically, by the presence of a high p_T lepton. The long lifetime appears as a displaced origin for tracks attached to the jet, which is detected by one of two types of algorithm. The first type attempts to explicitly reconstruct this secondary vertex (SV). The second type of algorithm uses the impact parameter (IP) of tracks, specifically its significance (IP/ϵ_{IP}), to indicate the consistency of a set of tracks with the PV. To calculate the impact parameter we take the tangent to a track at its point of closest approach to the PV and propagate this back to find the closest perpendicular distance to the PV. The algorithm also considers the point at which the back-propagated track would cross the jet axis. A real track must originate either at the PV or from a SV in the jet at some positive distance along the jet axis. For tracks which cross the jet axis on the wrong side of the PV we give the IP a negative sign. These negatively signed tracks cannot come from real SVs and so their spread gives a measure of the vertex resolution. For each of the track based *b*-taggers a ‘negative tag’ can also be formed using the tracks with negative signed IP, this is used in our estimation of the fake tag rate. Figure 2.7 is a simple graphical representation of a jet containing a *b*-quark and illustrates the characteristic features used by lifetime *b*-taggers.

Jet Lifetime Impact Parameter (JLIP) algorithm [42] The resolution of the tracking is used to calculate the probability for each track to belong to the PV. The probabilities are combined to give an overall probability that all tracks in the jet are consistent with the PV. A jet is tagged if it has a $JLIP_{prob}$ less than a cut value which can be optimised for the particular analysis scenario. Cuts range from ‘Very Tight’ (< 0.002) to ‘Super Loose’ (< 0.04).

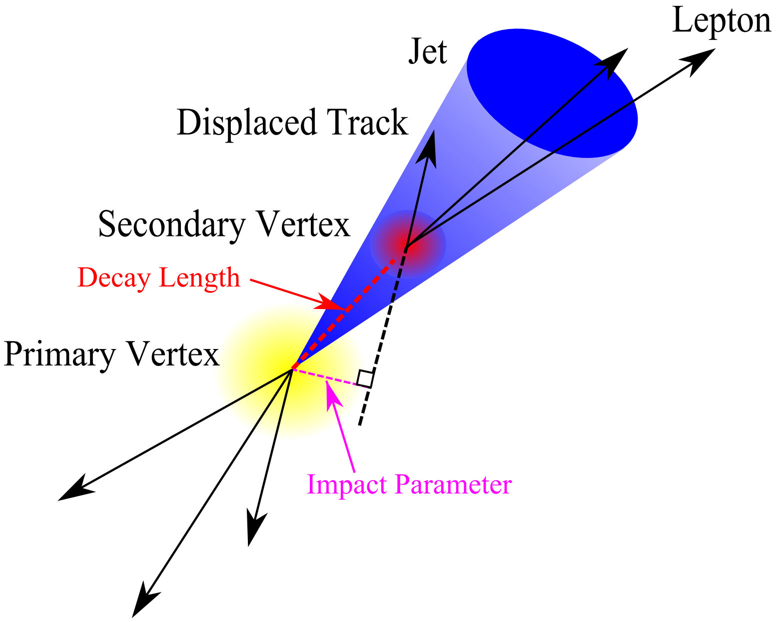

Figure 2.7: Schematic representation of a *b*-jet, illustrating the impact parameter and secondary vertex.

The probability is sensitive to single tracks with very high IP, potentially producing fake tags from a single mismeasured track. Such tracks could be complete fakes or real tracks mistakenly assigned to the jet. To protect against this a reduced probability, calculated with the highest significance track removed, is also available.

Counting Signed Impact Parameter (CSIP) algorithm [43] This tagger counts the tracks with large IP significance. If at least three tracks have a significance greater than two or two have a significance greater than three, the jet is tagged.

Secondary Vertex Tagger (SVT) [44] Tracks which are displaced from the primary are used to generate all possible secondary vertices (SV). A jet is tagged if a SV is found within $\Delta R < 0.5$ with a large enough decay length significance (DLS). The minimum DLS required is varied to give a variety of operating points, from tight (> 7) to loose (> 5).

Soft Lepton Tagger (SLT) [45] The jet is tagged if a muon is matched to a track attached to the jet and has more than a minimum $p_{T_{Rel}}$. The $p_{T_{Rel}}$ is the component of the muon's p_T perpendicular to the jet axis, where the jet axis is calculated including the muon track.

Chapter 3

Level-3 *b*-tagging

3.1 Introduction

The identification of b-quarks is an important tool in the analysis of a wide range of physics signals, from top physics to Higgs searches. Where a process of interest with b-quarks is kinematically similar to a light-flavour QCD background the background process is typically several orders of magnitude larger than the signal. Effective b-tagging can help reduce such background to a manageable level. Efficient b-tagging triggers should maximise the number of events with b-jets remaining in the samples whilst reducing the dominant light-jet backgrounds. The online algorithms are simpler versions of those used offline working at high efficiency operating points, albeit with higher fake rates. When optimising trigger tools the maximum rate of events which can be recorded and the limited processing time must be considered.

All b-taggers use either the b-quark's long lifetime or its muonic decays. Muons have a low enough rate to have high efficiency taggers of their own. However triggering on muons exclusively would only record a limited fraction of b-jets. Two lifetime taggers have been developed for Level-3 and this chapter deals with studies and commissioning of these tools. The ability to form a lifetime-tag for b-jets relies on the relational properties of the associated tracks, jets and vertices. Since b-tagging relies on the combination of information from multiple parts of the detector it requires multiple physics tools to be run which in total takes considerable time. The tracking algorithm consumes the most processing time.

3.1.1 Chapter Contents

The remainder of this introduction gives an overview of the Level-3 physics tools required to run the b-tagging algorithms and a brief summary of the challenges of

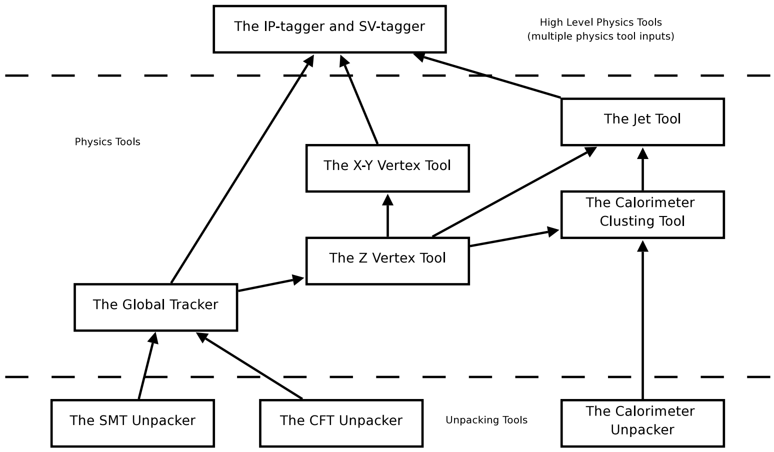

Figure 3.1: Schematic showing the tools required to run the Level-3 *b*-tagging filters.

developing Level-3 tools. Section 3.2 introduces the data, MC samples and offline software tools used for the Level-3 studies. Section 3.3 covers improvements to the IP based Level-3 *b*-tagger including a project to extend the track information used to three dimensions. The commissioning of a new secondary vertex *b*-tagging tool is described in Section 3.4. Section 3.3 is solely the authors work, while the final section was completed in collaboration with another PhD student.

3.1.2 Level-3 *b*-tagging Walk Through

The *b*-tagging tools are high level trigger tools, processing physics objects returned by other tools rather than handling raw data. Several of the following studies concern the effect on the *b*-tagging of changes to the lower level tools. Each of these tools is configurable via a set of parameters which may be tuned for specific analysis requirements and as conditions of luminosity and detector performance change. To clarify the roles and relationships of these tools the following section consists of an overview of the relevant tools. Figure 3.1 is a schematic showing how the Level-3 *b*-tagging tools depend on the jet, vertex and tracking tools, how these tools are related and how they depend on the unpacking tools which deal with the raw detector read-out.

3.1.2.1 An Overview of Level-3 Track Finding

The unpacking tools for the SMT [46, 47] and CFT [48] use the raw output of the SMT and CFT sub-detectors to construct ‘clusters’. A cluster represents the point where a charged particle has passed through a detector layer. Level-3 clusters are constructed in the x-y (axial) and z (stereo) detector layers.

The Global Tracker [49, 50] uses the clusters produced by the unpackers to reconstruct tracks. The time available for tracking at Level-3 is ~ 100 ms per event [32]. The Level-3 tracking algorithm must therefore be considerably simpler than the offline one, as a result it is less efficient and less pure although it is around one hundred times faster [51].

Axial tracks are found first starting with all pairs of hits in the outer two layers of the CFT and working inwards. As the track candidates are propagated inwards hits are added as long as the inclusion of the hit does not increase the track fit χ^2 by more than 10. Tracks are rejected if they are below a p_T threshold or if they are missing hits in more than one consecutive layer. Finally if any tracks share more than two hits the lower quality track is discarded.

Stereo tracking is performed using a fast histogramming approach [52]. To describe any stereo track two parameters are needed; Z_0, the z-coordinate at the closest approach to the origin, and $\tan(\lambda)$, the pitch of the track helix given by dZ/dS_{xy} where S_{xy} is distance along the track arc in the x-y plain. Each stereo hit should be consistent with a set of potential tracks described by a line in the $Z_0 - \tan(\lambda)$ plane. The point where these lines intersect for a set of hits gives the track parameters for the track made up of those hits. A 2D histogram is filled with these parameters and the most populous bin is used as a first approximation to the stereo track parameters. A linear fit is then used to find the optimal stereo track parameters.

3.1.2.2 An Overview of Level-3 Primary Vertex Finding

The z-vertex is determined by a histogramming method [53]. Tracks are binned in 1cm intervals between ±100cm based on z_0, the z-position of closest approach of the track helix to the origin in the x-y plane. The position of the z-vertex is calculated using the tracks within the two adjacent bins with the highest values for a weighted combination of track multiplicity and track p_T. The p_T weighting helps reject vertices from minimum bias events which will also be present.

Tracks which are consistent with the z-vertex are then used to find the x-y vertex which is found using an impact parameter (IP) minimisation [54]. The impact parameter is the closest perpendicular distance between the track arc and the vertex hypothesis. Tracks are approximated by straight lines, tangents at the point of closest approach to the beamspot in the x-y plane. The nominal beamspot is taken to be the average position of primary vertices from the previous data taking run and is parametrised as a function of z. The quantity to be minimised in determining the best vertex is:

$$\chi^2_{vertex} = \sum_{i}^{Ntracks} \frac{(d_0^i)^2}{(\sigma_0^i)^2} + \sum_{j=x,y} \frac{(b_j - V_j)^2}{(\sigma_j)^2} \tag{3.1}$$

where d_0^i are the impact parameters and the error on the IP is σ_0^i. The second summation constrains the vertex to be close to the beamspot position as it represents the most likely interaction point. The proposed vertex coordinates are $V_{x,y}$ and the beamspot position is $b_{x,y}$. The width, $\sigma_{x,y}$, is estimated from the spread of the vertices and is assigned to be 30μm. For the more accurate offline primary vertex finding tool this spread is measured to be 27μm.

3.1.2.3 An Overview of Level-3 Jet Finding

Jets found at Level-3 are similar to offline jets, although the algorithm is somewhat simplified. There is no jet energy scale correction at Level-3 and so a Level-3 jet will tend to have lower p_T than its offline counterpart. The Level-3 calorimeter clustering tool [55] passes clusters to the Level-3 jet tool [56] which forms $\Delta R \leq 0.5$ cone jets.

3.1.2.4 The *b*-tagging Tools

The *b*-tagging tools share a common system for matching tracks to each jet and to the PV. A track is attached to the nearest jet if the angle between them is less than 45°. This requirement is equivalent to a cone of $\Delta R < 0.7$, larger than the standard 0.5 cone used for track-jet matching offline. Any isolated tracks are discarded since both *b*-tagging algorithms make reference to the associated jet axis. Tracks are then required to pass the following 'good track' selection:

- At least 5 hits in the axial tracking detectors.
- A χ^2 for the axial track fit smaller than 5.
- At least 5 hits in the stereo tracking detectors.
- A χ^2 for the stereo track fit smaller than 5.
- The error on the z0 track parameter must be smaller than 0.4 mm.
- The track-vertex distance in z must be smaller than 2 cm.

Jets which have two or more 'good' tracks attached are called 'good' jets. Once the set of tracks for each jet has been created the algorithms for the two taggers diverge. The older impact parameter based tagger (IP-tagger) will be described in Section 3.3 and Section 3.4 will give details of the newly commissioned secondary vertex tagger (SV-tagger).

3.1.3 Challenges for Level-3

The Level-3 tools have undergone continuous development to maintain signal efficiency with increased background rejection as the instantaneous luminosities have increase during Run II. The rapidly increasing luminosity not only changes the desired working points but also affects input distributions through increased track and vertex multiplicities. Increases in track multiplicity in particular will have a significant effect on tool timing. Timing profiles must be monitored and tools re-tuned to account for the changing conditions.

In addition to the gradual drift of the beamspot during normal running, large movements in the beamspot occur during the annual Tevatron maintenance shutdowns. A large asymmetry in the position of the beamspot away from the coordinate origin gives a directional bias to tracking quality, since the sub-detectors nearer the PV have an effectively denser configuration of channels in $r - \phi$ than those opposite. It has proved very difficult for the Tevatron to correct such movements and so changes to the track clustering thresholds are preferred to recover efficiency. The Level-3 tools must be adaptable or robust against such changes to the beamspot position.

As the detector ages the increased number of dead channels will potentially have an impact on the Level-3 tools, as will any changes in the detector geometry during maintenance. There is also considerable effort required to ensure compatibility with the online framework and with offline tools and data formats, this activity is not described here but is left implicit.

3.2 Data, MC Samples and Tools

The following sections introduce the MC and data samples used in this chapter. The following studies were carried out using an offline trigger simulation framework called trigsim [57]. The standard data and MC samples used by physics analyses are reduced to standard reconstructed physics objects, to keep file sizes manageable the raw detector readout is removed. For studies of trigger tools the raw detector readout is necessary and so these standard samples cannot be used, alternative samples are made containing only the raw read-out. For studies comparing the Level-3 and offline results it was necessary to run separate algorithms over separate data file types for the same events in parallel.

	Run Number	Instantaneous Luminosity ($cm^{-2}s^{-1}$)	Number of Events
Low	188676	0.1×10^{32}	589495
Medium	206003	$0.6 - 0.7 \times 10^{32}$	309324
High	204905	$1.0 - 1.1 \times 10^{32}$	61684

Table 3.1: Data samples used.

3.2.1 MC Samples

Monte Carlo samples (MC) of $Z^0 \rightarrow b\bar{b}$ and QCD are used for studies of the Level-3 b-tagging tools. These samples were produced using Pythia [58] version 'v6 319'. A jet p_T cut of 12 GeV and a z vertex within ±35cm are required to mimic requirements in the *b*-tagging filter. Due to the complicated environment of a hadron collider significant data/MC differences occur. Data jets have a softer p_T spectrum, are fewer in number and contain fewer tracks. Track errors are also not accurately reproduced. As a result *b*-tagging will always appear to be more efficient in MC than in data. Thus significant data based tests are also required.

3.2.2 Data Samples

Studies using data avoid the above problems but are less straight-forward since the true flavour of data jets is unknown. For the purposes of these studies it is sufficient to make a *b*-enhanced and a standard data sample. Absolute efficiencies will not be measured, but comparisons of performance curves between tool versions remain valid.

To study the effects of changing luminosity, in Section 3.3.4, three samples were used as detailed in Table 3.1. In each case the complete run is used as a standard 'background' sample and the subset of events with an offline secondary vertex tag is used as a *b*-enhanced 'signal' sample. For the calibration of the IP-tagger 'negative tags' are used to generate a data sample with very low signal content as described in Section 3.3.2.

3.2.2.1 The *b*-enhanced Sample

The 'signal' samples produced from single data runs mentioned in the previous section are extremely limited in size due to the low production cross-section of *b*-jets. These samples are sufficient for studies mainly concerned with the fake rate. Section 3.4.4 deals with such a study on the effect of changing trigger conditions

Flavour	Initial fraction %	Tagging efficiency %	Fraction after two tags %
b	2.5	50	93.1
c	4.5	10	6.7
$udsg$	93	0.4	0.2

Table 3.2: Typical sample composition, the offline tagging efficiency and the resulting expected composition of the b-enhanced sample. The initial composition is that given by Pythia.

and luminosity on the fake rate of the IP-tagger. The errors resulting from the small signal samples became the limiting factor in this comparison of the tool performance for different scenarios. For the subsequent studies which concern the performance of the SV- and z-IP-taggers a larger b-enhanced sample was necessary.

The mechanics of the process used to generate the large b-enhanced sample are somewhat complicated due to the sparse population of such events across many runs, and therefore across many data files. Offline code can only be run on reconstructed events while online code must run on raw data, forcing the sample to be generated in a two pass process. Event lists were generated by running the offline secondary vertex tool over a large muonic data skim of eighty million events and requiring two jets with displaced secondary vertex tags. Assuming a typical flavour composition calculated from QCD MC allows us to calculate the composition after requiring two secondary vertex tags at the given working point. Table 3.2 gives the flavour composition from MC, the tagging efficiencies on the constituent samples and the resulting composition after two tags demonstrating that the b-enhanced sample is relatively pure heavy flavour. Only events with exactly two jets were used to avoid contamination of the b-enhanced sample by additional light jets.

The files containing each individual event are then accessed from central storage and the raw data for these events is concatenated. Due to the scarcity of such events each event is typically recorded on a different data storage tape. Finally we need to develop an infrastructure for running over events from disparate runs. During live running parameters which change on a run by run basis, such as the beamspot and solenoid polarity, are retrieved from the online database. In our standard trigger simulation framework these are generally supplied by hand as studies generally consider only a small number of runs. With the run number changing after every few events an automatic update method is necessary. Such an addition to the framework was developed to update the beamspot and extended to cover changes in the polarity, allowing such a large b-enhanced sample to be used for the first time in trigger studies at DØ. The code used to create this sample is still in use and has been used to increase the sample size with new data.

3.3 Studies with the Impact Parameter Tagger

3.3.1 Introduction

At the time of the work described in this chapter the IP-tagger was the only b-tagging trigger tool running online. The tool was the only b-tagger available during Run IIa and as such was an important component of the triggers used to collect the majority of the data events used in the analysis described in Chapter 5.

3.3.2 The IP-tagger Algorithm

The matching of tracks to jets and to the PV is described in Section 3.1.2.4 and is the same for all the Level-3 b-taggers. Once the sample of tracks associated with a jet is found the IP is calculated for each track. Since tracking precedes PV finding, as seen in Figure 3.1, the tracking tool parametrises tracks in terms of the absolute fixed origin in the x-y plane, independent of the choice of PV. The IP calculation is simplified if the PV is used as the coordinate origin, so all track parameters are first recalculated in this frame of reference. The distance of closest approach (DCA) for a Level-3 track is its closest distance to the coordinate origin. Here the DCA is recalculated so that it gives the closest approach to the PV. The error on the calculated DCA is given by the tracking tool, this error propagates into the errors on all the distances we calculate starting from the DCA, such as the IP. At the scale of the DCA the curvature of tracks is very slight (radius of curvature, $R \gg$ DCA). Since the distances under consideration are well within the beam pipe, scattering and energy loss through interactions are very low. We can thus approximate the track as an arc in the x-y plane with a constant velocity in z. The tracking resolution is very different in the x-y plane ($\sim$15 μm) compared to the z-direction ($\sim$ 0.5 mm). For this reason the existing implementation of the IP-tagger calculates IPs in the x-y plane (Section 3.3.5 describes a study into the feasibility of including the z-information in the IP calculation).

The IP is measured at the most likely origin of the track, i.e. its point of closest approach to the jet axis. This point is either where the track crosses the jet axis or where it is parallel to it. To find this point we step back along the track arc in the direction which minimises the distance between the track and the jet axis. Figure 3.2 shows some of the parameters of the iterative position finding, where R is the radius of curvature, dS is the distance moved along the path, $\vec{D}$ is a tangent to the current point and (Xc,Yc) is the centre of the track arc. In reality the curvature is far less than that used for illustration in the figure. The IP will change little since

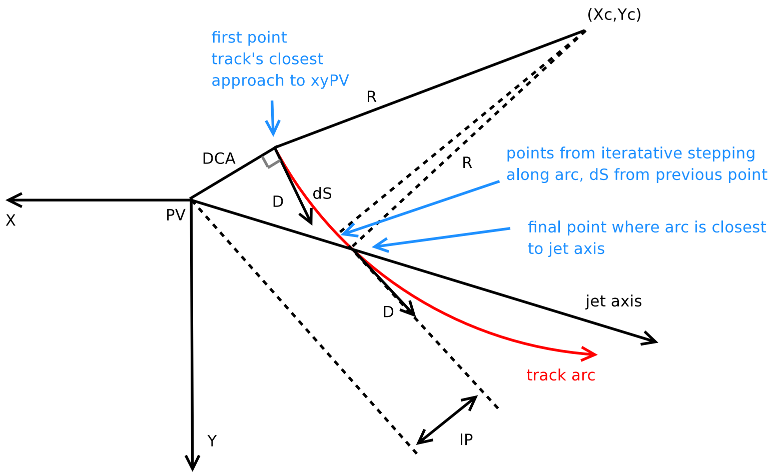

Figure 3.2: Schematic to relate the Level-3 track parameters and the iterative IP finding algorithm.

the tangential vector $\vec{D}$ changes little. The IP is the distance from the PV to the final point, perpendicular to the track tangent $\vec{D}$.

At this stage tracks are discarded if the calculated track parameters do not fulfill the IP-tagger track selection:

- Impact parameter < 0.25 cm
- Decay length < 1 cm
- Track-jet distance < 0.007 cm
- Track-jet distance significance < 10

where the decay length is the distance along the jet axis to the point where the track and jet are closest, and the track-jet distance is the separation of track and jet at this same point.

Around 50% of tracks in a background sample will cross the jet axis on the opposite side of the PV from the jet itself. Such a position is clearly inconsistent with a physical origin in the jet; such an IP can only be explained by the finite resolution of the PV and tracking. To represent this information each IP is given a sign indicating which tracks cross the jet axis downstream (+) and upstream (-) of the PV. The negatively signed track sample contains very few tracks from long-lived particles and therefore is not used to find b-jets. These tracks are not used directly by the online tool. However this lifetime-free track sample is ideally suitable for calculating the resolution of the tracking system. The resulting function

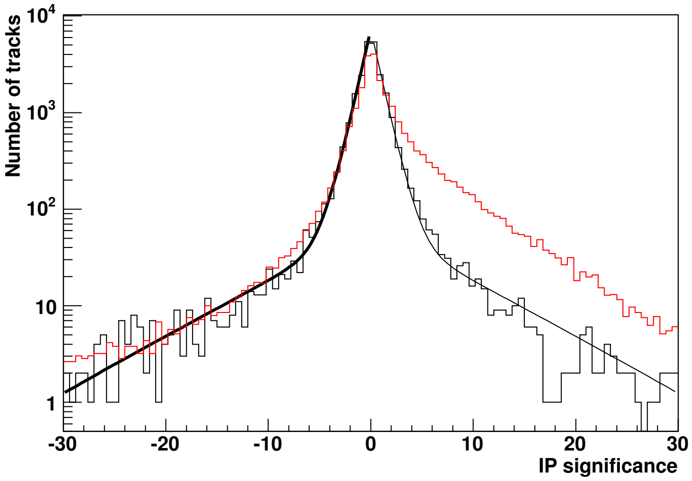

Figure 3.3: The signed IP significance for tracks in the medium luminosity data run (black) and the *b*-enhanced data sample (red). The resolution function produced from a fit to the negative signed IPs and reflected in the y-axis is also shown (black curve). An excess over the resolution function can be seen for positive IPs in the *b*-enhanced sample.

is then used by the IP-tagger to separate signal lifetime from resolution effects in the positive signed track sample as described below.

The IP significance for a track, IP_{sig}^{track}, is given by the track's IP divided by the error on the IP. A resolution function is produced from a fit to the IP significance distribution of the negative signed tracks. This fit function is reflected in the y-axis to produce the resolution function for positive IP significances. Figure 3.3 shows the distribution of signed IP significance in two data samples, the medium instantaneous luminosity background sample and the *b*-enhanced sample. The resolution function is also shown, it is in good agreement with the negative signed IP significance distribution and clearly shows the excess of positive signed IP significance for the *b*-enhanced sample.

The resolution function is normalised to unity and integrated above each track's IP significance to give the probability for the track to have a larger absolute IP significance, known as the 'track probability' P_{track}:

$$P_{track} = \frac{\int_{IP_{sig}^{track}}^{\infty} f_{res}}{\int_{-\infty}^{\infty} f_{res}} \tag{3.2}$$

where f_{res} is the resolution function. The smaller the track probability the more unusual the track's IP and the less likely the track really comes from the PV. Figure 3.4 shows the track probability distributions for the medium instantaneous luminosity background data sample and the *b*-enhanced sample. Tracks with negative signed IPs are shown in the top-left plot to give an idea of the distribution of probabilities for tracks whose displacement is due to resolution effects. The top-right plot in Figure 3.4 shows the track probabilities for the positively signed tracks. The bottom plot shows the excess of positive signed tracks over negative signed tracks. The total 'event probability' is found by taking the product of the track probabilities for all positive signed tracks with a combinatoric normalisation factor which accounts for the increased chance of finding some high IP tracks in events with higher multiplicities. The complete equation to calculate the event probability is given in Equation 3.3 where $\mathcal{P}^{i}_{track}$ are the track probabilities and N_{trks} is the number of tracks. Since Π is a probability it is always between zero and one, therefore $-ln(\Pi)$ is always a positive number.

$$\mathcal{P}_{event} = \Pi \times \sum_{j=0}^{N_{trks}-1} \frac{(-ln(\Pi))^{j}}{j!} \text{ with } \Pi = \prod_{i=1}^{N_{trks}} \mathcal{P}^{i}_{track}(IP_{sig} > 0) \tag{3.3}$$

This probability $\mathcal{P}_{event}$ is known as 'the b-tag' and is the primary output of the tool. It is the probability that all tracks in the event are consistent with the primary vertex. A b-tag of zero therefore strongly suggests the presence of one or more b-jets. The distribution of b-tags for the medium instantaneous luminosity background data sample and the b-enhanced sample are shown in Figure 3.5. A jet based tag is calculated in the same way, but is not used in the existing trigger.

3.3.3 IP Resolution Function Refits

3.3.3.1 Introduction

The track probability is calculated using the resolution function determined from a fit to the distribution of negative IP significances. This function will vary over time due to changes to the characteristics of the detector and tracking algorithms, increases in luminosity and detector occupancy, detector ageing, changing trigger thresholds and trigger tool improvements. As these changes are continuously occurring new resolution functions are produced after each major revision of the trigger algorithms. The IP-tagger was first available in the trigger version known as v13 which was used online from June 2004. The studies in this section and the rest of this chapter concern the v14 version which followed in June 2005 and ran online until the end of Run IIa in March 2006. Figure 3.6 shows the difference between the tracks' IP and IP-significance for the same set of events for the two trigger versions, it is clear that a fit to the v13 IP significance will not accurately describe the distribution in v14.

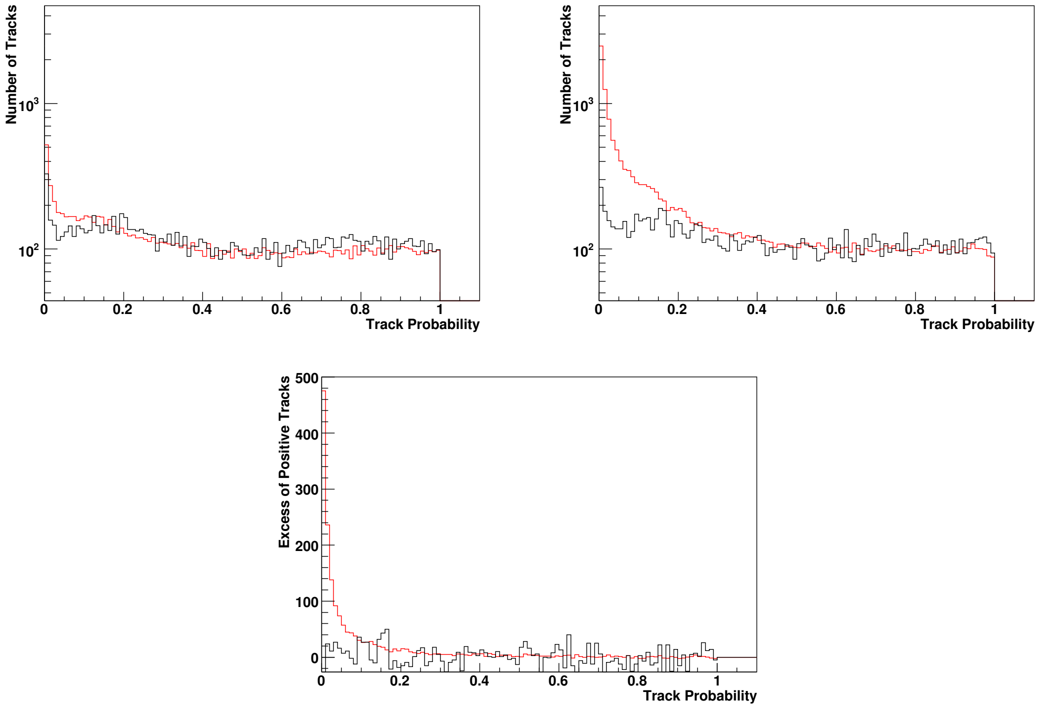

Figure 3.4: The track probability in the medium luminosity data run (black) and the b-enhanced data sample (red). Upper left: the track probabilities for tracks with negative signed IP, such tracks are not used in the event probability calculation but give an indication of the fake rate due to resolution. Upper right: the track probabilities for tracks with positive signed IP, these are used to calculate the event probability. Lower: the excess of positive signed tracks formed by subtracting the negative signed tracks' probability distribution from the positive signed tracks' distribution. All three figures are normalised so that the number of tracks with negative probability are the same in the b-enhanced and background samples.

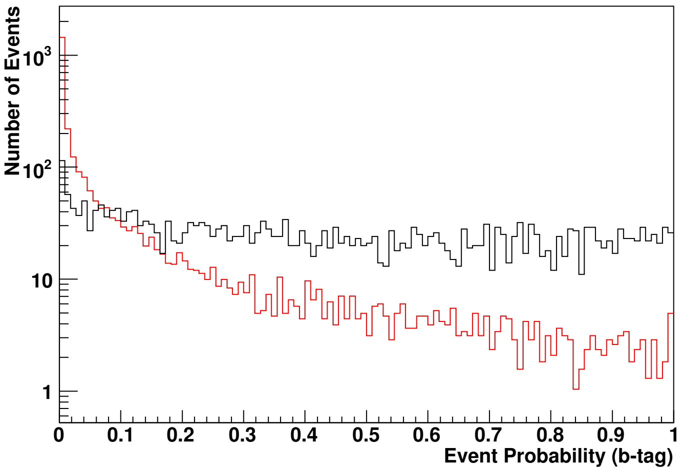

Figure 3.5: The event probability (b-tag) in the medium luminosity data run (black) and the b-enhanced data sample (red). Good separation between signal and background is apparent. The two samples are normalised to have the same number of events passing the b-tagger preselection.

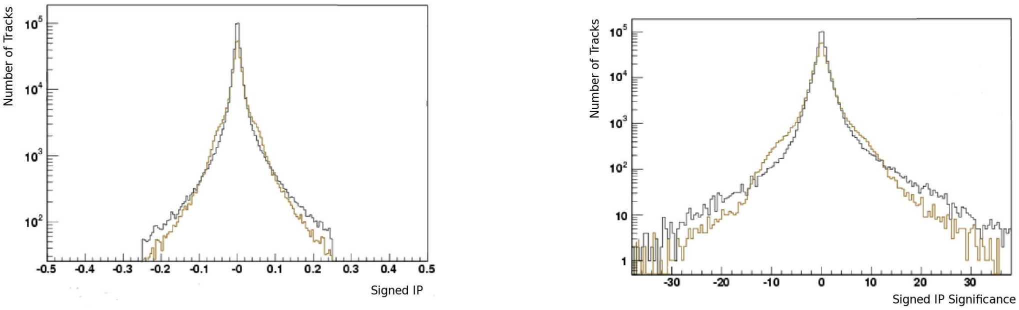

Figure 3.6: The signed IP (left) and signed IP significance (right) distributions for the v13 (red) and v14 (black) trigger versions for a low luminosity run. More tracks pass the b-tagging track selection in v14, to allow the shapes to be compared the histograms in these figures are normalised to equal area.

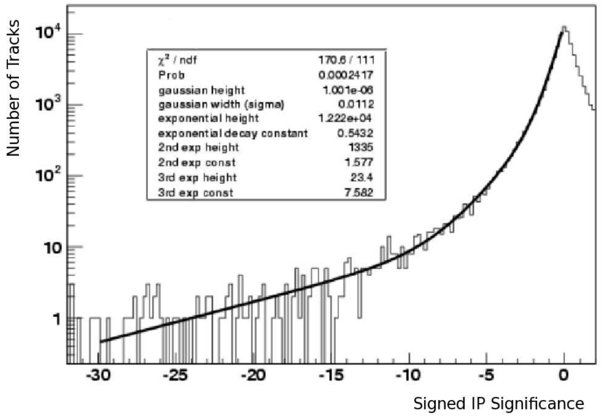

Figure 3.7: The chosen best fit to the resolution function with three exponentials. The χ^2_{dof} for the fit is 1.54.

3.3.3.2 Resolution Function Refits and Tests

The pre-existing resolution function fit consisted of a Gaussian and two exponentials to account for the long tails. This function was generated using a small single run sample by the tool authors at the time the tool was first used online. The function gives a reasonable approximation to the distribution seen in v13, but to exploit the full potential of the tool the fit should accurately represent the data it will be used upon. The availability of more data since the original fit should also allow a better description of the tails.

Several functional forms were investigated, beginning with the form used for v13. All possible fit functions using between zero and two Gaussians and between one and three exponentials were tried, along with a fit to four exponentials. The quality of the fits was judged by their χ^2 per degree of freedom, χ^2_{dof}. The IP resolution function was derived from the low luminosity sample and tested on the medium luminosity sample. At the time of the refit the low luminosity sample represented a typical run and the medium luminosity sample was one of the highest instantaneous luminosity samples available. The fit using three exponentials was found to best represent the distribution and is shown in Figure 3.7.

3.3.3.3 Conclusion

Retuning of the IP-tagger resolution function is essential as the luminosity increases, tracking algorithms change and the detector ages. The refit described here was a better quality fit to a larger data sample than the original function and was necessary to maintain the performance of the tool. The fit function derived ran online until the end of Run IIa and the tools developed have been used in subsequent refits. The suitability of the derived resolution function for the medium and high luminosity samples is discussed in the following section.

3.3.4 Stability of the IP-tagger with Luminosity and Tool Changes

3.3.4.1 Introduction

Changes to the tracking and vertex finding tools will potentially affect the number and quality of tracks available to the IP-tagger. As the luminosity rises the optimal configuration for each of these tools will change. During the lifetime of the v14 resolution function discussed previously, a number of major changes to these tools occurred, along with an order of magnitude increase in the peak instantaneous luminosity. In this section two potential improvements are investigated. In the first scenario the p_T threshold for tracks used to reconstruct the PV is raised. In the second scenario, along with this change, a retuned SMT unpacker is introduced. Both these scenarios were tested on the low, medium and high instantaneous luminosity samples. The purpose of the studies was not to motivate the changes, which were suggested by other studies, but to confirm that they might be implemented without damaging the efficiency of the IP-tagger.

Raising the PV track p_T threshold reduces the CPU time spent on Level-3 tracking, which is correlated with the instantaneous luminosity. The time overhead of low p_T tracking becomes more significant as the increasing luminosity leads to increasing numbers of simultaneous 'minimum-bias' interactions superimposed on top of the hard scatter. The IP-tagger still requires tracking to be carried out down to low p_T (1 GeV). The PV is found for every event while b-tagging is only performed on a small fraction of events. Overall it is more economical to run tracking to a higher threshold (3 GeV) on all events and rerun it (down to 1 GeV) on the small subset of events for which an IP-tagger trigger term is called and which pass the IP-tagger preselection cuts. As instantaneous luminosity increases this change will be needed to help keep the tracking algorithm within the timing constraint.

The improvements to the SMT unpacker included 'hot' chip removal[1], pedestal retuning and splitting/merging of clusters. The result was a $\sim 10\%$ increase in the number of SMT clusters in 'good' axial tracks and a $\sim 70\%$ increase for stereo tracks.

---

[1] The offline tracking group identified SMT read-out chips with abnormally high occupancy ($> 25\%$) as 'hot'. Clusters from these chips account for a third of clusters in an event on average and are mostly fake. Such clusters should be removed.

3.3.4.2 Effect on Track DCA

Figure 3.8 shows the track DCA significance for tracks with negative signed IP. Since almost all real tracks (all those not from the decay of long lived particles such as b-quarks) physically originate in the PV the distribution of DCA's is close to a Gaussian around zero with a width reflecting the errors of the tracking system. If the track errors are correctly assigned the distribution of track DCA over the DCA error, the DCA significance, will have a width of 1. This width is referred to as the DCA "pull". The measured pull for the existing Level-3 configuration and track errors is a little greater than one, suggesting a slight underestimate of the track errors or of the beamspot width parameters, $\sigma_{x,y}$ introduced in Section 3.1.2.2. The purpose of this study is not to correct the DCA pull but to try to anticipate potential future problems for Level-3 b-tagging due to changes in the Level-3 tools and in the instantaneous luminosity. The pull is not significantly affected by the trigger list changes, suggesting that the raised PV track p_T threshold is not adversely effecting the quality of the PV. The higher instantaneous luminosity runs show a slight increase in the DCA pull. This is as expected since PV reconstruction is less accurate in a high track multiplicity and high multiple interaction environment. To correct the DCA pull it will become necessary to refit the DCA errors, but the size of the effect is currently small.

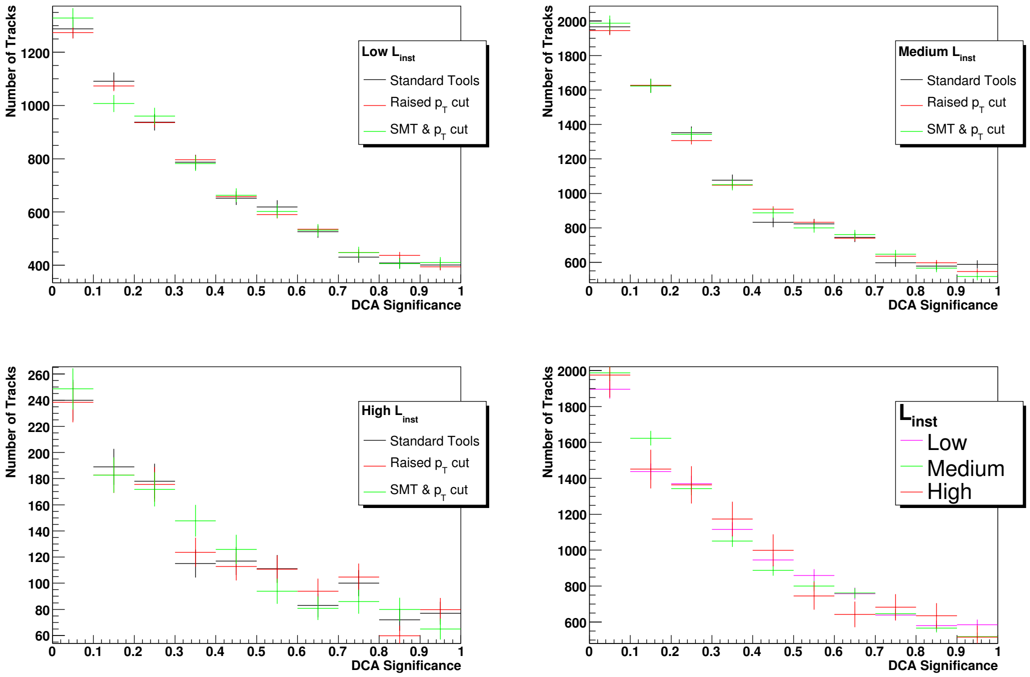

Figure 3.8: The track DCA significance for negatively signed tracks in the low (top left), medium (top right) and high (bottom left) instantaneous luminosity data runs. Three track samples are shown for each luminosity, produced using the standard tools (black), with a raised PV track p_T threshold (red) and with both the SMT unpacker improvements and raised PV p_T threshold (green). The lower right plot shows a direct comparison between the track distributions in the low, medium and high luminosity runs for the scenario with the upgraded SMT unpacker and raised PV p_T threshold.

3.3.4.3 Effect on the Number of Tracks

Figure 3.9, shows the number of tracks associated with each jet. The effect of the trigger list changes is small, but there is significant degradation at increased luminosity. The higher luminosity samples have a higher average jet multiplicity but an increasing fraction of these jets are not associated with tracks. This suggests either that the method of assigning tracks to jets is failing, that there is an increase in fake jets or that there is a reduction in tracking efficiency in the medium and high instantaneous luminosity runs, and that the problem increases with instantaneous luminosity. Any one of these effects has the potential to affect the tagger's performance. The most significant change with increasing luminosity seems to be a large increase in the number of jets with no tracks associated, these are most probably noise jets. As long as the overall number of jets with associated tracks does not decrease the efficiency of the IP-tagger should be maintained.

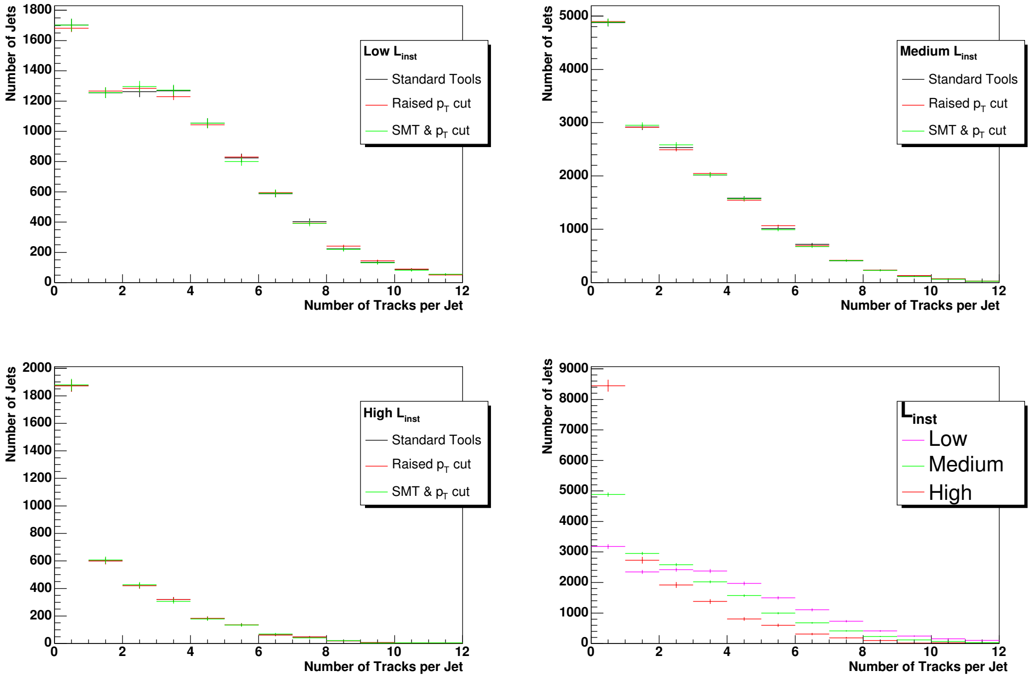

Figure 3.9: Number of tracks per jet for the low (top left), medium (top right) and high (bottom left) instantaneous luminosity data runs. Three jet samples are shown for each luminosity, produced using the standard tools (black), with a raised PV track p_T threshold (red) and with both the SMT unpacker improvements and raised PV p_T threshold (green). The lower right plot shows a direct comparison between the track distributions in the low, medium and high instantaneous luminosity runs for the scenario with the upgraded SMT unpacker and raised PV p_T threshold.

3.3.4.4 Effect on the Number of Tracks Used in the PV

Figure 3.10 shows the number of tracks in the PV for the three different scenarios for the low, medium and high instantaneous luminosity data runs. As expected raising the p_T threshold on tracks used to calculate the PV reduces the number of tracks and the number of PVs found. This difference is much less significant for PVs with a large number of associated tracks. Figure 3.11 shows the effect of the parameter changes on the medium luminosity sample, for all events (as in Figure 3.10) and for events which have at least two jets suitable for online b-tagging (jets with at least two tracks passing the IP-tagger track selection). The second sample is more like the sample of events for which the IP-tagger will be run in the trigger. After applying this requirement the effect of raising the PV track p_T threshold is much less significant. The PVs which are affected by the raised threshold seem to mainly be badly reconstructed PVs, unsuitable for b-tagging.

As the instantaneous luminosity increases the mean number of tracks in the average PV decreases, most significantly for the high instantaneous luminosity sample. This reflects the expected increasing difficulty of vertexing and tracking at the higher instantaneous luminosities.

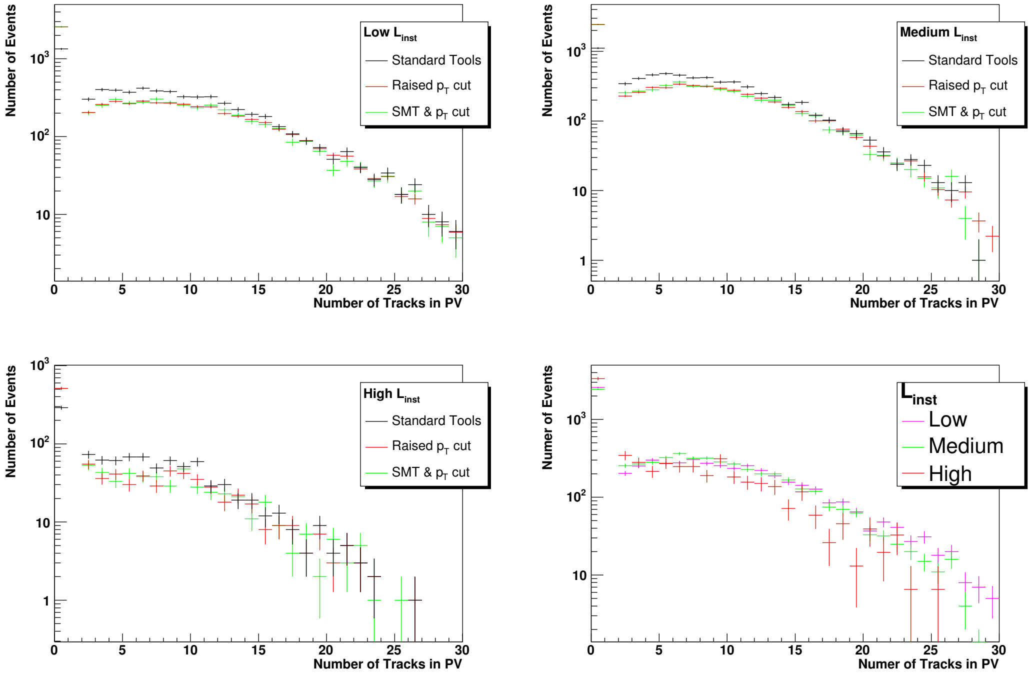

Figure 3.10: Number of tracks used to build the PV for the low (top left), medium (top right) and high (bottom left) instantaneous luminosity data runs. Three event samples are shown for each luminosity, produced using the standard tools (black), with a raised PV track p_T threshold (red) and with both the SMT unpacker improvements and raised PV p_T threshold (green). The lower right plot shows a direct comparison between the track distributions in the low, medium and high instantaneous luminosity runs for the scenario with the upgraded SMT unpacker and raised PV p_T threshold.

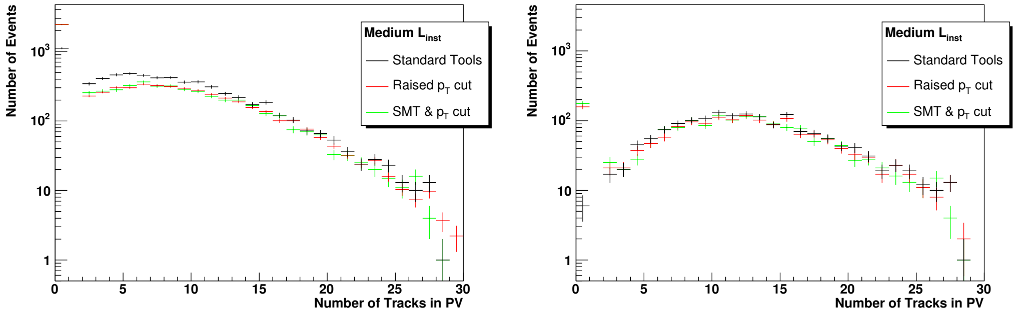

Figure 3.11: Number of tracks used to build the PV for the medium luminosity run, in all events (left) and in events with at least two jets containing at least two 'good' tracks (right).

3.3.4.5 Effect on the IP-tagger Track Probability

Figure 3.12 shows the track probability for negatively signed tracks. The resolution function used in this version of IP-tagger was derived on the low instantaneous luminosity sample as described in Section 3.4.3. This version of the IP-tagger was required to function routinely at luminosities equivalent to the medium luminosity sample and up to luminosities represented by the high luminosity sample, although at the time the resolution function was derived no instantaneous luminosity comparable to that of the high luminosity sample had been recorded.

A flat track probability for these tracks indicates that the resolution function parametrisation accurately describes the track resolution of the sample. The tool changes do not have a significant effect on the track probability, suggesting the resolution function fit remains an accurate reflection of the track resolutions for the tested scenarios. The improved SMT unpacking parameters do not majorly improve the track resolution. The departure from a flat track probability at small probabilities indicates the difficulty in describing the tails of the IP significance distribution in the resolution function. These low track probabilities correspond to large IP significance tracks. The problem of underestimating the resolution function tails and producing an excess of low probability tracks is seen at all three luminosities. In the low luminosity sample the high probability region is quite flat and this flatness degrades at the higher instantaneous luminosities.

The majority of real b-jets will have tracks with probabilities in the lowest bin, these are the most powerful tracks in the event probability combination. The increase in the number of tracks seen in the first bin due to resolution effects, in the negatively signed sample seen here, will be mirrored in the positively signed sample and will lead to an increase in the fake rate and a reduction in the power of the tool. This degradation of the overall performance will require a new resolution function fit as the luminosity increases further.

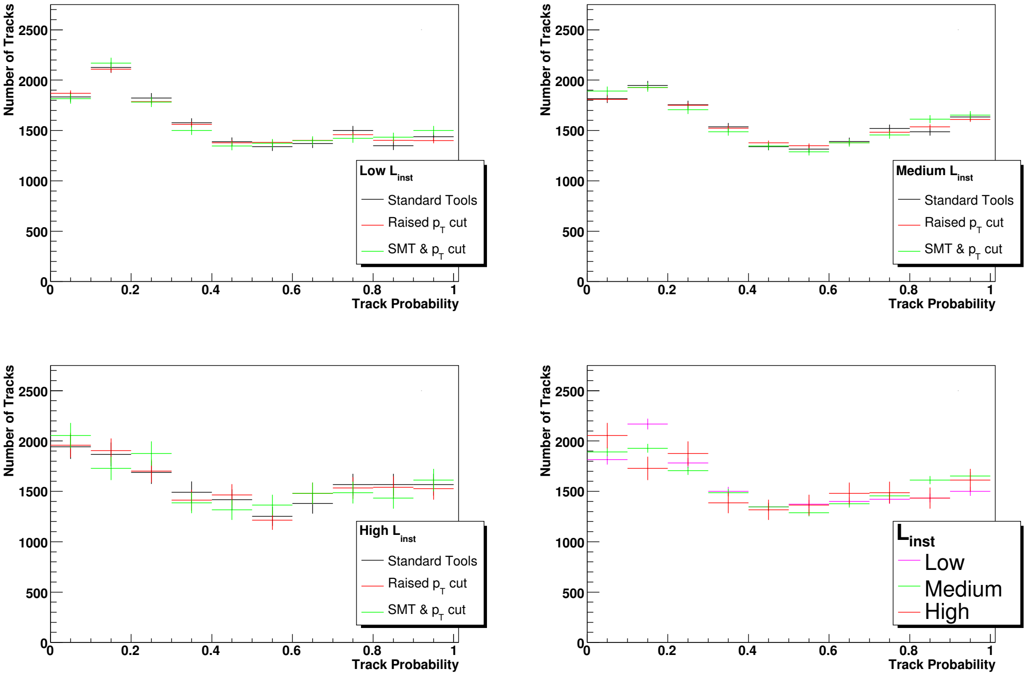

Figure 3.12: The track probability distribution for negatively signed tracks in the low (top left), medium (top right) and high (bottom left) instantaneous luminosity data runs. Three track samples are shown for each luminosity, produced using the standard tools (black), with a raised PV track p_T threshold (red) and with both the SMT unpacker improvements and raised PV p_T threshold (green). The lower right plot shows a direct comparison between the track distributions in the low, medium and high instantaneous luminosity runs for the scenario with the upgraded SMT unpacker and raised PV p_T threshold.

3.3.4.6 Effect of Changes on the IP-tagger Performance

Performance curves were created for the three data runs with the three tool configurations. Signal events were identified by requiring a tight SVT tag offline. Events containing b-jets are rare and so the total number of events in the signal sample for a single run is small. The resulting statistical errors limit our ability to compare performance on single runs.

Figure 3.13 shows performance curves for the three tool configurations for the medium luminosity run. It is not possible to resolve a significant change in performance between the scenarios, within errors they are equivalent. This is consistent with the minimal effect seen on the track and PV parameters for the set of 'good' tracks and vertices used by the IP-tagger. The performance on the higher luminosity sample is similarly not affected by the change in scenario, however the performance for all three scenarios is reduced compared to the medium luminosity sample. The signal statistics are poor for the high luminosity sample and so a comparison is made at the $\sim 80\%$ efficiency nominal performance benchmark of the IP-tagger. At this working point the fake rate at medium luminosity of $32 \pm 5\%$ increases, as expected, to $44 \pm 8\%$ for the high luminosity run. Here efficiency is defined with respect to the total number of data events with 'good' Level-3 jets and an offline SV tag. The fake rate was measured with respect to the total number of events with 'good' jets in each full data run at each luminosity. The decreased performance at high luminosities gives increased motivation to the commissioning of a second b-tagging tool described in Section 3.4 and for the retuning of the IP-tagger once high instantaneous luminosity becomes common place.

To overcome the issue of limited signal statistics the large b-enhanced sample described in Section 3.2.2.1 was created and was used for the remaining studies in this chapter.

3.3.4.7 Conclusion

It was confirmed that the proposed changes to the trigger tools which feed information to the IP-tagger cause only a limited reduction in the IP-tagger's performance. These changes are necessary to stay within processing constraints in the more demanding high instantaneous luminosity environment. The latest versions of the trigger use these new trigger list parameters. The IP-tagger was tested on the highest instantaneous luminosity run available. The algorithm continued to be capable of high efficiency but suffered a $\sim 40\%$ relative increase in fake rate.

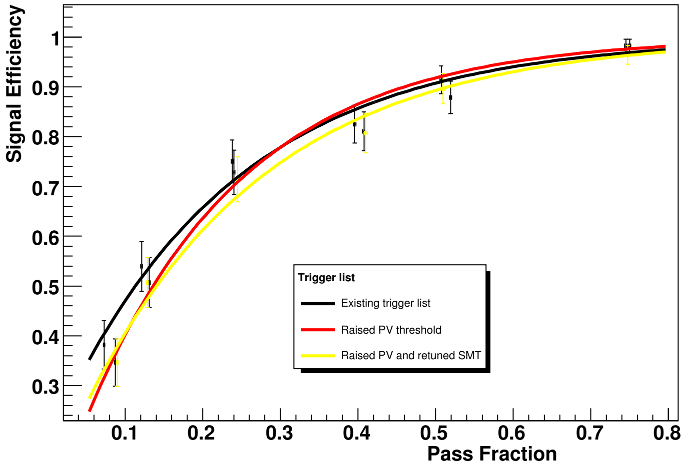

Figure 3.13: Performance for three tool configuration scenarios for the medium luminosity data sample. Results are shown for the existing trigger list (black), the raised PV threshold list (red) and the combined PV threshold and SMT fix list (yellow).

3.3.5 Z-dimension Impact Parameter Tagging

3.3.5.1 Introduction

In the existing tool the z-component of the IP is neglected, in fact the whole process of finding the closest approach to the jet is performed in two dimensions only. One of the significant effects of the improvements to the SMT unpacker mentioned in the previous section was an improvement in the number of stereo hits per track. With more stereo SMT hits per track there is more accurate information for tracking in z and the z-resolution has consequently improved. This improvement makes it possible to consider using the z-component of the IP. In this section the development of an extension to the IP-tagger is described which would calculate the z-component of the IP (the z-IP) and include this information in the final tool output.

3.3.5.2 Tracking Errors

Despite the improvement to the z-tracking and the reduction in the mean error, the errors in z are still much larger than those in x-y. The resolution of the PV in the z-direction has also improved, but it should be remembered that because of the geometry of the silicon detector the z-resolution will always be larger than that in x-y.

Recalculating the DCA in three dimensions and looking at the DCA pull in z suggested that the z-errors provided by the tracker were underestimated as the pull was greater than one. The z-errors from the tracker had not been updated since their initial implementation and so new estimates were developed for use in the z-IP study. These were produced by binning the tracks in the medium luminosity run by number of SMT hits and coarsely by the quality of the PV as indicated by the number of tracks used to build the PV. Figures 3.15, 3.16 and 3.14 show the fits for tracks with varying numbers of SMT hits for PVs with five or fewer tracks, six to ten tracks and more than ten tracks. A fit to the DCA distribution of the negative signed tracks was made with a pair of Gaussians for each track and PV category. The width of the dominant Gaussian was taken to be an improved estimate of the error. The non Gaussian peak in the distributions for DCA values close to zero reduces the quality of the fits and makes these Gaussian errors conservative.

3.3.5.3 Implementing the tool

The IP algorithm has four components where two dimensions are specifically used:

- Reduction of track to a track arc.

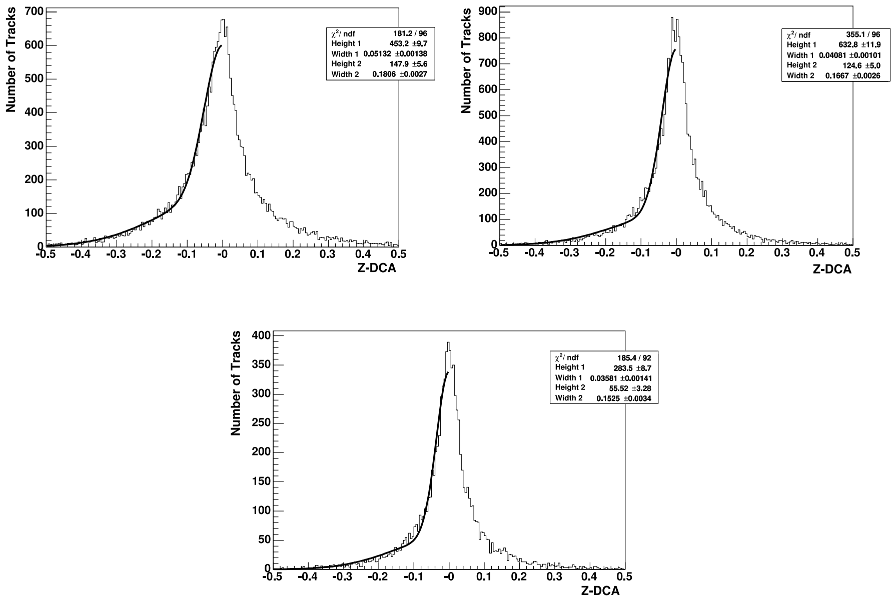

Figure 3.14: The fits to z-DCA (in cm) for tracks from events with five or fewer tracks used to reconstruct the PV. The tracks are divided into samples with 2 (top left), 3 (top right) and more than 4 SMT hits (bottom). The resulting Gaussian widths give the new z-track errors for the z-IP tool.

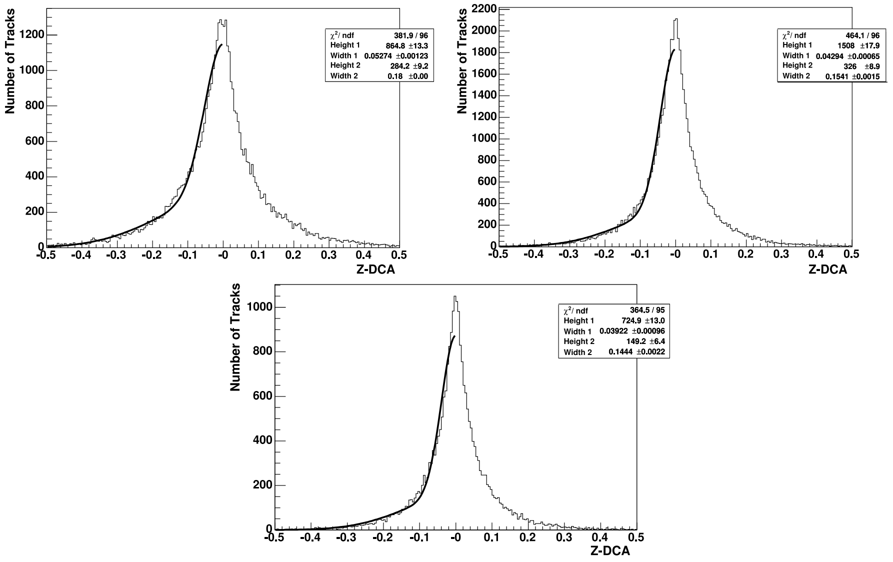

Figure 3.15: The fits to z-DCA (in cm) for tracks from events with between six and ten tracks used to reconstruct the PV. The tracks are divided into samples with 2 (top left), 3 (top right) and more than 4 SMT hits (bottom). The resulting Gaussian widths give the new z-track errors for the z-IP tool.

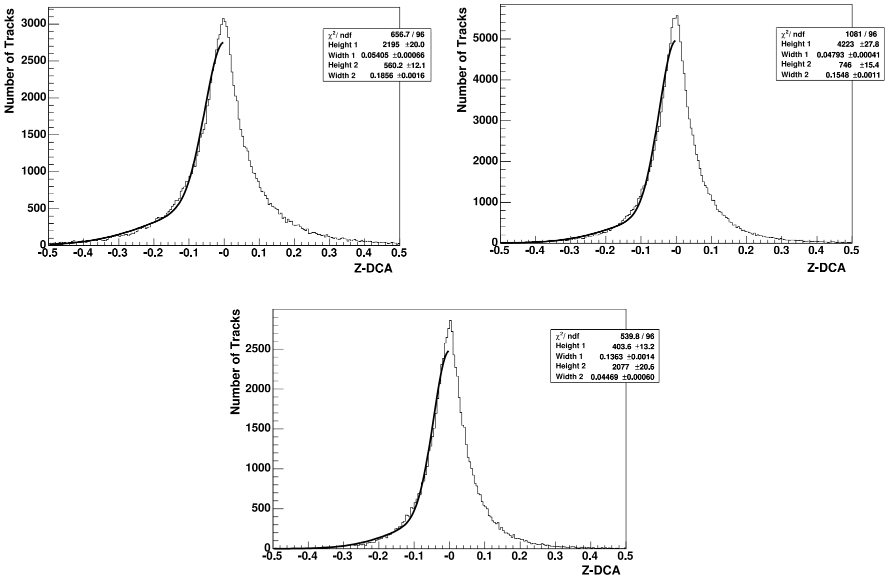

Figure 3.16: The fits to z-DCA (in cm) for tracks from events with more than ten tracks used to reconstruct the PV. The tracks are divided into samples with 2 (top left), 3 (top right) and more than 4 SMT hits (bottom). The resulting Gaussian widths give the new z-track errors for the z-IP tool.

- Finding the closest approach of the track to the associated jet, the DCA.
- Giving the track a sign based on its origin upstream (-) or downstream (+) of the PV.
- Calculating the IP.

The 2D track arc can be replaced with the required 3D equivalent, a track helix, by including a constant velocity in the z-direction. The existing 2D algorithm uses only the x and y components of the track position vector and so is not affected by this change to the z-component. The track arc is replaced with the track helix throughout the shared code base for the existing and z-IP extended taggers.

When searching on the track helix for the DCA to the jet, two points exist which could be considered optimal. Firstly the point which simply gives the closest approach in 3D. Secondly the point which gives the closest approach if the separation in each dimension is weighted with its relative error. In the first approach the z-component of the track-jet separation tends to dominate, the point with smallest z-separation is usually chosen irrespective of the x-y coordinates. This result seems unnatural since the x-y tracking is more accurate, as indicated by the smaller errors. The second approach behaves as might be expected, finding very similar points for calculating the DCA in the 2 and 3D cases. Since the stereo errors are an order of magnitude larger than the axial ones when using this approach the 2D point is always a good approximation of the 3D point. The second approach is used since it takes into account the relative reliability of tracking in different dimensions. To maximise code reuse and minimise the overhead for running the z-IP tool the point found using the 2D closest approach is used with the 3D track helix to produce the 3D DCA. In this scenario the sign given to the track for the 3D tool should be the same as the sign in the 2D tool.

The IP is calculated by a vector method which could easily accommodate the z-information. However because of the large differences between axial and stereo tracking the resolution functions for the 2D and the z-IP significance are produced separately. The probabilities can then be combined to form a 3D probability per track. In order to maintain the independence of the 2D tagger it was decided instead to calculate a separate x-y and z event probability to be combined with separate cuts in the actual trigger.

3.3.5.4 Track Selection for the z-IP tool

As the stereo track parameters are subject to larger errors, the track selection was loosened to:

- Z impact parameter < 2 cm.
- Z decay length < 5 cm.
- Z track-jet distance < 5 cm.
- Z track-jet distance significance < 30.

These parameters were optimised for peak z-IP tagger efficiency (considering the efficiency of the z-IP only, not in combination with the IP-tagger). The best performance was found using the track-jet distance significance as the main track selector. The decay length and track-jet distance cuts are loosened to the point where they reject very few tracks. The minimum 'good track' preselection defined in Section 3.1.2.4 was also required.

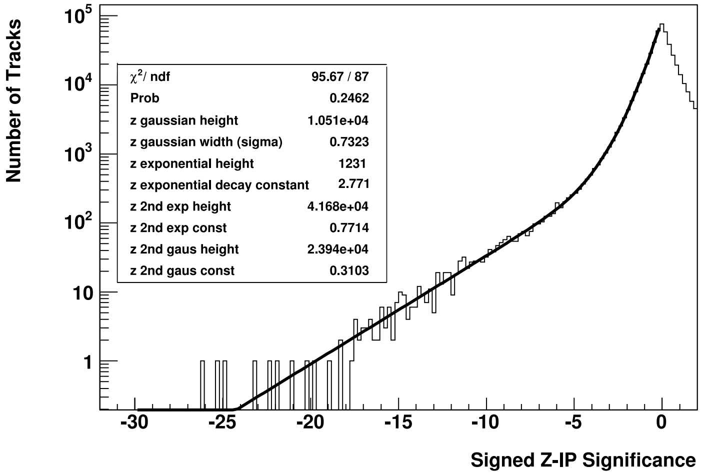

Figure 3.17: The fit to the z-IP resolution function using two Gaussians and two exponentials.

3.3.5.5 Measuring the Resolution Function

The optimal resolution function was determined in a similar way to the 2D function. In this case, the best fit function was two Gaussians with two exponentials. As before only the negatively signed tracks are used to measure the resolution function. The sign given by the 2D tagger is more reliable than the sign from the z-IP tool. Comparing the sign given to each track in the 2D and z-IP tools indicates that for heavy flavour jets the signs are more correlated than for light jets. The 2D sign can therefore be used to confirm the sign derived by the z-IP algorithm, whilst rejecting mostly background. Using only tracks given the same sign by both z-IP and 2D taggers was seen to give better performance than using either sign alone. Only these tracks with like signs are used by the final z-IP tagger. The z-IP significance is shown in Figure 3.17. A comparison of the z-resolution function in Figure 3.17 to the x-y function in Figure 3.7 shows that the z-track errors introduced above are reasonable since the spread in IP significances are comparable.

Figure 3.18 shows the resolution function for the z-IP compared to the IP significance distributions for both the *b*-enhanced and background data samples for positively and negatively signed tracks. The fact that there is some excess over the resolution function for positive IPs in the signal sample suggests that there is some useful *b*-tagging information in the z-IP for at least some tracks. The excess is considerably smaller than the corresponding one for the 2D tagger.

Figure 3.19 shows the track probability for the *b*-enhanced and background data samples. The track probability for background tracks is relatively flat. This flatness

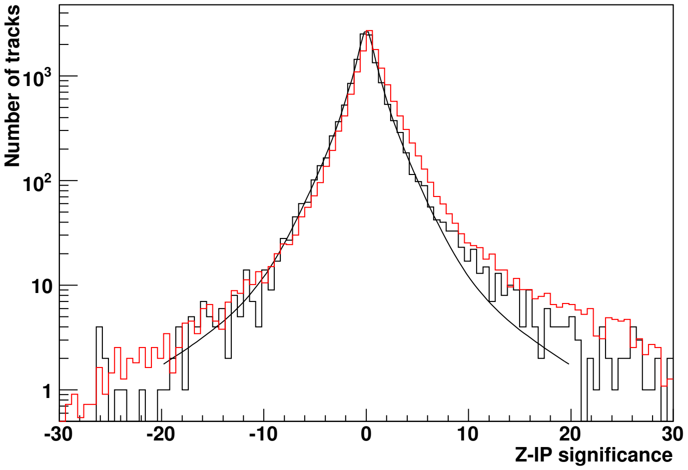

Figure 3.18: The signed z-IP significance for tracks in the background sample (black) and *b*-enhanced sample (red). The resolution function derived from a fit to the negative signed IPs is also shown for comparison, a small excess over the resolution spread due to real long-lived particles can be seen especially in the signal sample.

indicates the quality of the resolution function fit. An excess of signal tracks is seen with positive track probabilities, this is required if the tool is to be able to discriminate. Comparing the excess of positive signed tracks in Figure 3.19 with that shown in Figure 3.4 for the 2D IP-tagger indicates how much less favourable the lower resolution z-direction tracking is for IP *b*-tagging. With the 2D IP-tagger the excess of positive signed tracks is concentrated at low track probability, this is because the mean path of the *b*-quark is significantly larger than the width of the resolution function. For the z-IP the resolution width and mean path are of the same order. The *b*-quark mean path is still larger than the z-resolution, resulting in the broadly distributed excess seen in Figure 3.19. In the final event probability (Z *b*-tag), shown in Figure 3.20, a limited separation can be seen between signal (at lower probabilities) and background (at higher probabilities).

3.3.5.6 Performance and Conclusion

Using the large *b*-enhanced and background data samples the performance of the z-IP-tagger was measured. The signal efficiency was defined using the offline double SVT tight tagged signal sample:

$$\text{Signal Efficiency} = \frac{\text{Number of Level-3 tagged events from the signal sample}}{\text{Number of events in the signal sample}} \tag{3.4}$$

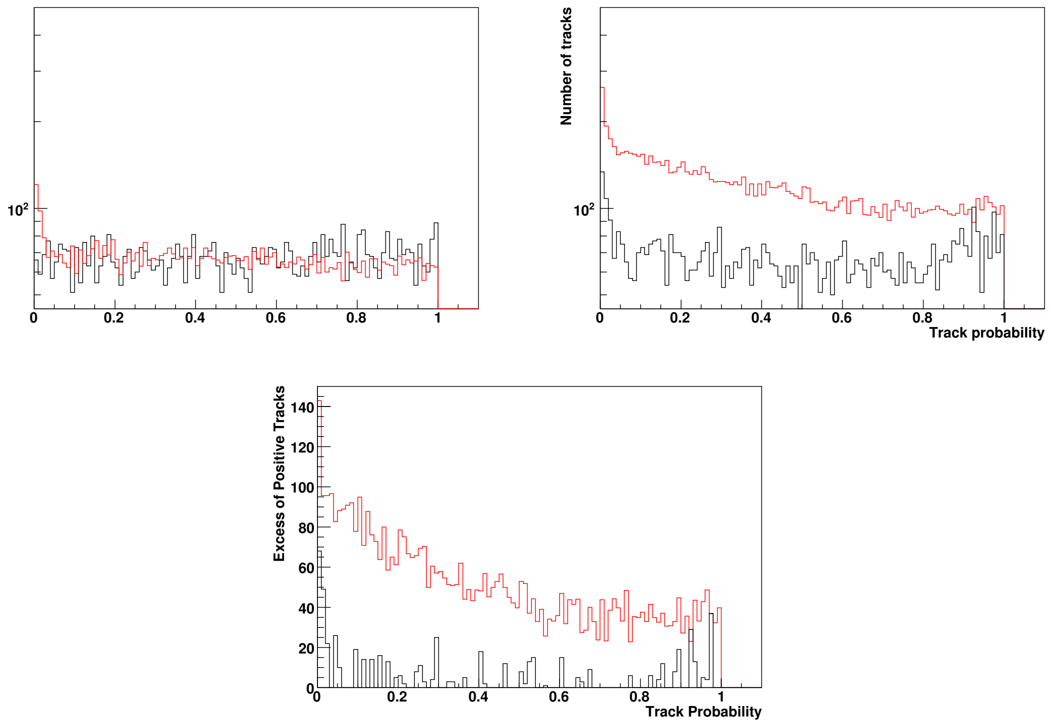

Figure 3.19: The z-IP track probability in the medium luminosity data run (black) and the b-enhanced data sample (red). In the top-left figure the track probabilities for tracks with negative signed z-IP are shown, such tracks are not used in the event probability calculation but give an indication of the fake rate due to resolution. The top-right figure shows the track probabilities for tracks with positive signed z-IP, these are used to calculate the event probability. The bottom figure shows the number of positive signed tracks with a given probability minus the number of negative signed tracks with the same probability. All three figures are normalised so that the number of tracks in the resolution sample, those with negative signed z-IP, are the same in the b-enhanced and background samples.

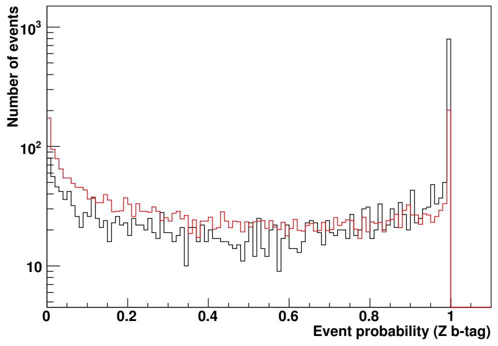

Figure 3.20: The event probability (Z-b-tag) for the z-IP tagger in data events for the medium luminosity background sample (black) and the b-enhanced sample (red).

A background sample consisting of the medium luminosity data run was used to calculate the rate:

$$\text{Rate} = \frac{\text{Number of Level-3 tagged events from the background sample}}{\text{Number of events in the background sample}} \tag{3.5}$$

The performance is shown in Figure 3.21. As expected z-IP tagger performs less effectively than the IP-tagger. Whilst the z-IP tagger does give rejection of $\sim 40\%$ for 80% efficiency this is considerably less than the $\sim 75\%$ rejection of the 2D IP-tagger for the same efficiency.

In a real trigger filter the z-IP tool would be combined with the 2D IP-tagger. Figure 3.22 shows the performance of a 'z-IP AND IP-tagger', a 'z-IP OR IP-tagger' and the unmodified IP-tagger without z information. To generate the curves for the combined taggers the cut on the weaker tagger, the z-IP, was fixed and the IP-tagger event probability cut varied. For the 'OR' a tight cut on the z-IP event probability of < 0.1 was required. For the 'AND' a loose cut of z-IP event probability < 0.9 was used. Whilst the z-IP tagger shows some ability to separate signal and background when used on its own, directly combining the tagger's results with those from the existing 2D IP-tagger does not result in significantly improved performance. It is possible that a more subtle combination of results from the two taggers could improve efficiency. Future plans for Level-3 include a multivariate *b*-tagger which could potentially combine the z-IP and IP-taggers with the secondary vertex tagger introduced in the following section using a neural network.

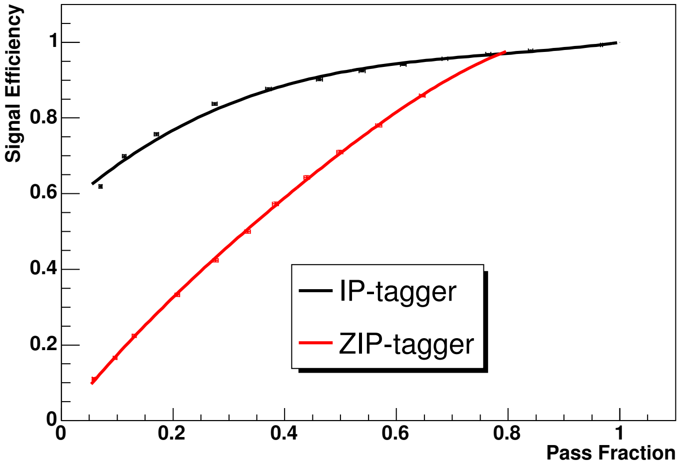

Figure 3.21: Performance curves for the z-IP and existing taggers run independently. Error bars are included but are small.

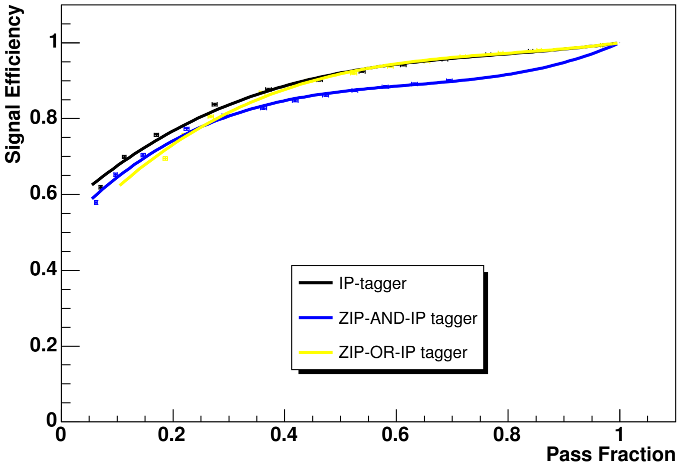

Figure 3.22: The performance of the 'z-IP AND IP' tagger (blue), the 'z-IP OR IP' (yellow) tagger and the existing 2D IP-tagger (black). Errors are shown but are small.

3.4 Studies with the Secondary Vertex Tagger

3.4.1 Introduction

The IP tagger considers whether the set of tracks associated with a jet are compatible with the PV. The tool does not consider the relationship between the tracks themselves as the probabilities for each individual track are simply combined. The decay of a b-quark inside a jet produces tracks radiating from a single point, the secondary vertex. Finding this point provides an additional way to reject tracks with erroneously high IP, which will have a different apparent origin. Such tracks may be fake tracks or real tracks associated with the wrong jet. Because the IP-tagger does not reconstruct the SV these tracks would not necessarily be rejected and constitute a source of fake tags for that tool.

The reconstruction of a SV requires at least two tracks from the b-decay, more than two such tracks may be required to build a useful SV if one track is mismeasured or if the SV is particularly close to the PV. As such the IP-tagger may be more powerful for lower track multiplicities than a SV based approach. Considering the complementary strengths and weaknesses of the two approaches has led to the development of both algorithms, both for the Level-3 trigger and for offline reconstruction. Combining the results from separately optimised lifetime taggers leads to an improvement in both efficiency and rejection over the individual taggers.

The secondary vertex tagger [59] was developed to explicitly reconstruct secondary vertex candidates in the jet. As with the IP-tagger the workings of the tool are confined to the x-y plane.

3.4.2 Tool Development

The secondary-vertex finding algorithm was developed from the IP minimisation method used for finding the PV as described in Section 3.1.2.2. This algorithm is considerably simpler than the offline secondary vertex finding algorithm described in Section 2.4.2, due to the time constraints at Level-3. A finite number of secondary vertex candidates are investigated at a set of 'evaluation points' (EP) along the jet axis. The EPs can be considered as a starting point for each minimisation, equivalent to the coordinate origin in the PV calculation. Candidate SVs are found from EPs up to a distance 0.9 cm from the PV along the jet axis spaced by 30μm (a total of 300 points). The tool considers tracks not used in the the PV and which pass certain quality requirements (p_T and χ^2). These tracks are linearised at their point of closest approach to the EP. The linearised tracks are used to calculate IP

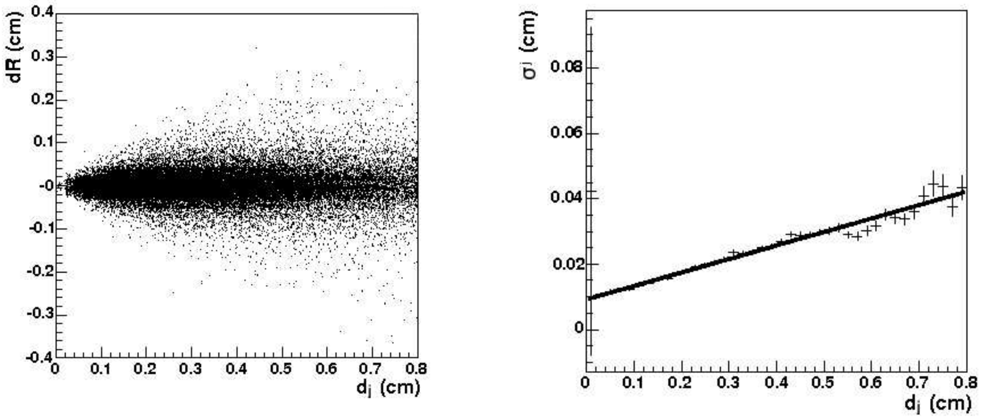

Figure 3.23: The distribution of secondary vertices found by the offline tool around the jet axis (left), the width of Gaussians fit to the distribution at a given dj (points, right) and the linear parametrisation of the width in terms of dj (line, right).

significances from the PV and the EP. Tracks are rejected if they have too small a significance from the PV or too large a significance from the EP.

If there are two or more tracks remaining the vertex candidate is found by minimising the χ^2 as in Equation 3.6.

$$\chi^2_{SV} = \sum_{t}^{N_{tracks}} \frac{(d_0^t)^2}{(\sigma^t_{d_0})^2} + \sum_{i=x,y} \frac{(j_i - V_i)^2}{(\sigma_i^j(d_j))^2} \tag{3.6}$$

where N_{tracks} is the number of tracks passing the selection cuts, V is the vertex candidate position, d_0 is the distance of closest approach from the track to the vertex candidate, j is the jet axis and σ^j is the expected spread of secondary vertices around the jet axis which is a function of the distance along the jet axis, d_j. The σ^j parameters were taken from Gaussian fits to the distribution of offline secondary vertices around the jet axis as a linear function of distance along the jet. The distribution of offline secondary vertices in ΔR and d_j can be seen on the left of Figure 3.23, where ΔR is the perpendicular distance of the SV from the jet axis. The width parameters from Gaussian fits to this distribution at a set of fixed d_j values and the linear parametrisation of these widths are shown on the right of Figure 3.23. The first term in Equation 3.6 is the contribution from the tracks and the second is a constraint to the jet axis, acting like the beamspot constraint in Equation 3.1.

The vertex candidate is initially produced from all available tracks. Tracks are then progressively removed and the χ^2_{dof} is recalculated. This method is known as ‘strip down’ as opposed to ‘build up’ vertexing. The χ^2_{dof} is calculated from the minimised total χ^2 above using the equation:

$$\chi^2_{dof} = \frac{\chi^2}{2 \times N_{tracks} - 2}. \tag{3.7}$$

If the track contributing the largest increase in the χ^2_{dof} adds more than a threshold $\delta\chi^2_{max}$ it is removed and the process is repeated. The vertex is stable once no track can be removed to reduce the χ^2_{dof} by more than the threshold or no more than two tracks remain.

Once all the possible EPs have been investigated the tool scans through the final χ^2_{dof}s along the jet axis, the EPs with locally minimum χ^2_{dof} are the vertex candidates produced by the tool. Where two adjacent minima are separated by a local maximum, the two SVs are not resolved unless the size of the peak-to-trough difference is more than a threshold value χ^2_{split}. If the minima are not separated in this way the candidate with the lower χ^2_{dof} is taken to be the correct vertex.

The output of the tool is the list of these minima for all jets, these are the SVs found for the event. Each SV has a decay length, a decay length significance (DLS), a number of tracks and a χ^2_{dof}. The filter script can use all of these variables, including the number of secondary vertices required in a jet or an event. For the following studies the most powerful variable, the DLS, is used as the discriminant to gauge the power of modifications to the tool.

3.4.3 Tool Commissioning

In the previous section there are a number of parameters which are not specified. These parameters either cut on tracks passed to the SV algorithm or control parts of the algorithm. A study to determine the optimum choice of these parameters is described in this section. In the order that they arise in the description the parameters are:

Track p_T - Minimum p_T for tracks to be included in the SV IP minimisation.

Track χ^2_{dof} - Maximum χ^2_{dof} of tracks.

Track PV significance - Minimum IP significance from the PV for tracks.

Track SV significance - Maximum IP significance from the current EP for tracks.

Maximum track χ^2_{dof} contribution - Maximum increase in SV candidate χ^2_{dof} when including a track.

Vertex candidate χ^2 separation - Minimum step up between two EPs with local minimal χ^2_{dof}.

Running the tool with a given parameter set will result in a number of proposed SVs in one or more jets. Filters using the tagger will require one or more SVs with a decay length significance (DLS) greater than a given cut. An 'operating point' for the SV-tagger consists of a parameter set and DLS cut. The set of operating points is defined to allow maximum efficiency for a range of rejection factors which may be needed for future trigger conditions. To study the optimal parameter set for the tagger we vary each of the above parameters in turn and compare the efficiency and rate for a number of benchmark DLS cuts.

The b-enhanced sample described in Section 3.2.2.1 is used to provide signal events. The fact that the sample is generated using the offline equivalent of the SV-tagger, the SVT defined in Section 2.4.2, means we will over estimate the efficiency, since we will be considering signal events which are known to be suitable for secondary vertex tagging. This bias is not a problem since the aim of the study is to find the best operating parameters for tagging such events, not to make an accurate measurement of the final efficiency for all b-jets. The signal sample efficiency is defined as:

$$\text{Signal efficiency} = \frac{\text{Number of Level-3 SV tagged signal events}}{\text{Number of signal events}}. \tag{3.8}$$

The low luminosity data sample introduced in Section 3.2.2 is used to calculate the rate:

$$\text{Pass Fraction} = \frac{\text{Number of Level-3 SV tagged background events}}{\text{Number of background events}}. \tag{3.9}$$

This rate is an estimate for the overall output rate of the tool running online since it is calculated using a sample of events from a single run without removing the small number of real heavy flavour events. This data run was recorded before the introduction of the IP-tagger into Level-3 and so the proportion of real b-jets is very small, and similar to the proportion for jets input to Level-3.

The optimal tool parameters are chosen so as to maximise efficiency, where two scenarios can achieve equal efficiency the scenario with the lower fake rate is chosen.

3.4.3.1 Track p_T

Only tracks with p_T above this cut are available for use in the SV. The p_T cut cannot go below 0.5 GeV since this is the minimum p_T track which the tracker reconstructs.

Raising the p_T threshold will improve the purity of the tracks (fewer fake or misconstructed tracks) but will reduce the efficiency. Figure 3.24 shows the performance of the SV-tagger for p_T thresholds between 0.5 and 1.5 GeV. In line with expectation the SV-tagger loses some power but becomes purer at higher p_T cuts. The highest peak efficiency is attained by using all tracks down to the minimum available from the tracker, therefore the cut of 0.5 GeV is chosen.

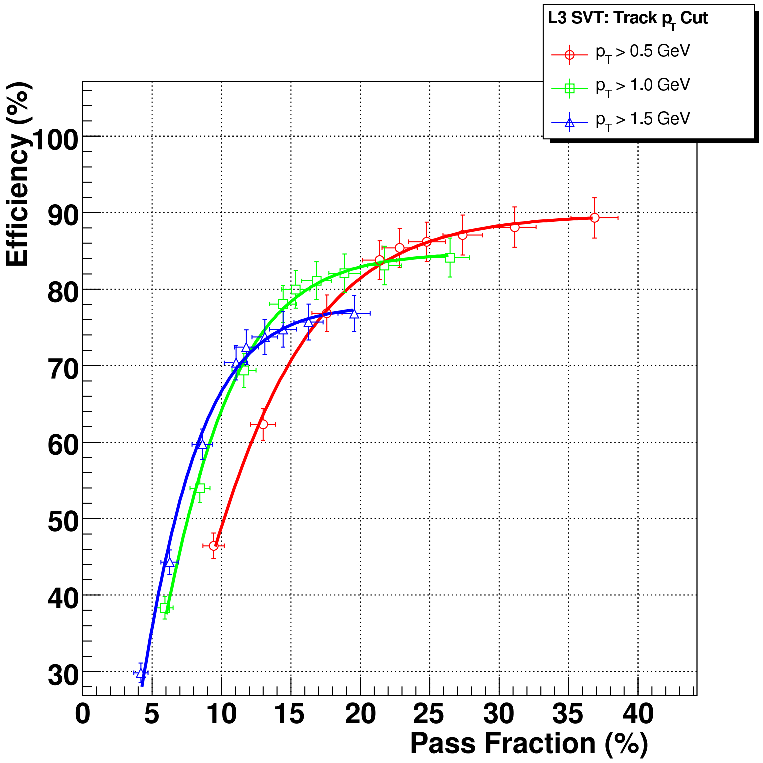

Figure 3.24: The performance of the SV-tagger using tracks with a minimum p_T of 0.5 GeV (red circles), 1.0 GeV (green squares) and 1.5 GeV (blue triangles). Statistical errors are included.

3.4.3.2 Track χ^2_{dof}

The χ^2_{dof} of a track is a significant indication of the track quality. A tight χ^2_{dof} cut will reduce the presence of fake and poorly reconstructed tracks, simultaneously losing some real tracks from the sample. Decreasing the cut value is therefore expected to have a similar effect to increasing the p_T cut.

Figure 3.25 shows the performance of the SV-tagger with a selection of χ^2_{dof} thresholds between 1.0 and 5.0. The vertexing algorithm will reject fake tracks itself and so is relatively robust when poorer quality tracks are allowed in. For the tightest scenario considered we begin to cut into the tagging efficiency, whilst also reducing the fake rate. While all but the tightest cut scenario are within errors equivalent the loosest cut is chosen since a slightly higher maximum efficiency is possible. The fact that the performance changes little between χ^2_{dof} cuts of 2 and 5 suggests that tracks with high χ^2_{dof} are either rare or are mostly rejected, without the need for this hard cut, within the iterative χ^2_{dof} minimisation process. The tuning of this parameter is therefore not expected to have a significant impact on the tool's performance and the loosest value, $\chi^2_{dof} < 5$, is chosen.

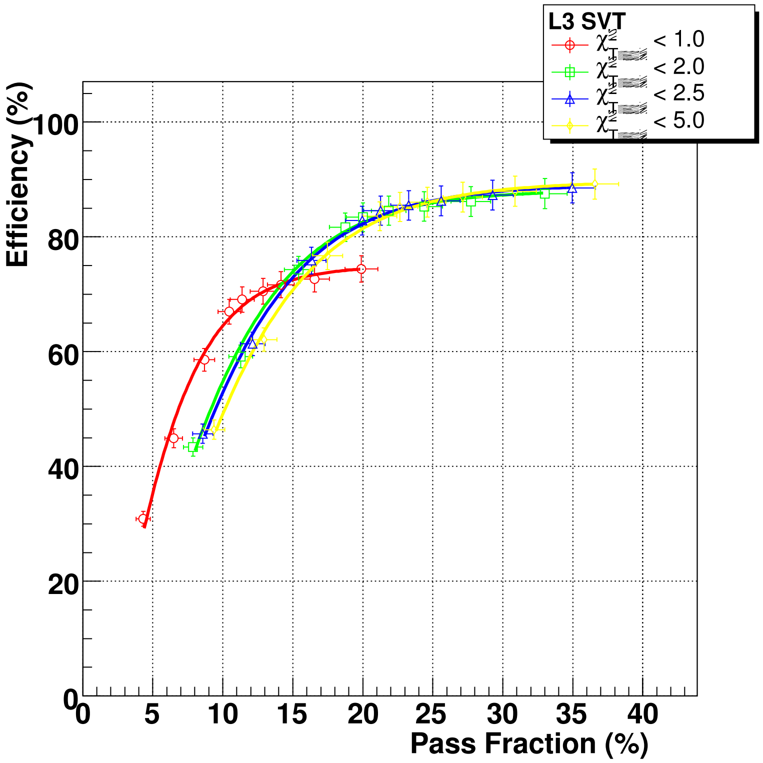

Figure 3.25: The performance of the SV-tagger using tracks with a maximum track fit χ^2_{dof} of 1 (red circles), 2 (green squares), 2.5 (blue triangles) and 5 (purple diamonds).

3.4.3.3 Track IP Significance From Primary Vertex

The IP significance with respect to the PV, IP$_{sig}^{PV}$, is calculated including both the track and PV errors. Tracks with a large IP$_{sig}^{PV}$ are more likely to come from a SV. Conversely tracks with low IP$_{sig}^{PV}$ are likely to originate from the PV. A minimum IP$_{sig}^{PV}$ cut can be used to prevent tracks consistent with the PV from being considered for inclusion in the SV.

The performance of the SV-tagger with a set of different IP$_{sig}^{PV}$ cuts between 0 and 5 is shown in Figure 3.26. This parameter has a very strong effect on the efficiency and purity of the tagger. As such the parameter is used to define a set of operating points for the tagger which give maximum efficiency for a broad range of rejection factors.

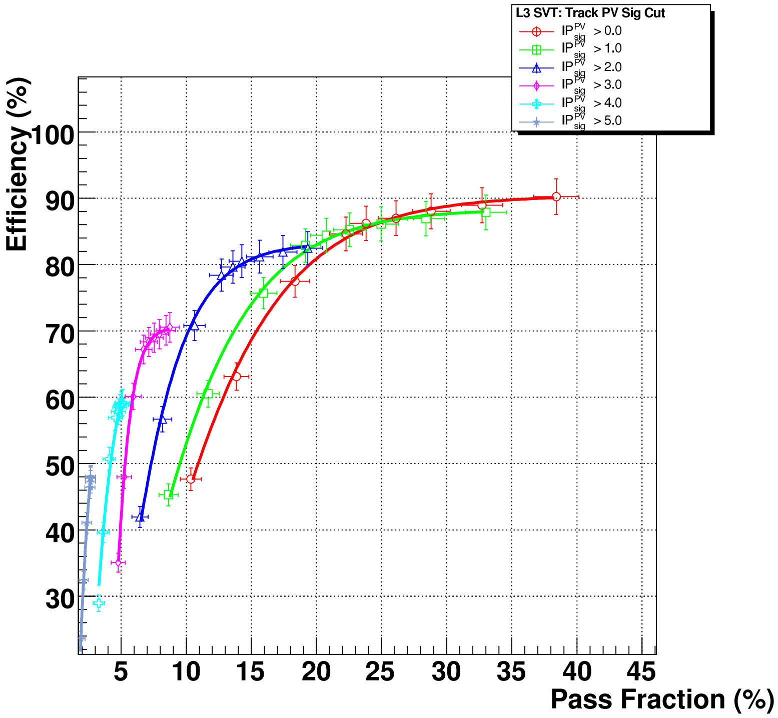

Figure 3.26: The performance of the SV-tagger using tracks with a minimum IP significance from the PV of 0 (red circles), 1 (green squares), 2 (blue triangles), 3 (purple diamonds), 4 (cyan crosses) and 5 (grey stars).

3.4.3.4 Track IP Significance From Secondary Vertex Candidate

For each EP at which candidate SVs are constructed tracks which are not compatible with the EP can be excluded from the minimisation. Such tracks may be mismeasured tracks from the SV or tracks not from the SV at all. Excluding these reduces the information available to construct the secondary vertices and could lead to a small number of signal events failing the minimum 2-jet condition. Including these tracks could pull the minimum χ^2 EP away from the true SV or result in more fake SV candidates being reconstructed. To remove such tracks an IP significance from the EP, IP_{sig}^{SV}, less than a given cut is required.

Figure 3.27 shows the performance curves for a set of taggers with different requirements for maximum IP_{sig}^{SV}. In the section of the performance curves with highest efficiencies the scenarios with no cut and with IP_{sig}^{SV} less than 6 or 8 are all within errors the same. A cut at 8 is chosen since it offers slightly higher maximum efficiency than a cut at 6. Although the performance with the loosest cut and with no cut are identical within errors, it is preferable to use a finite cut value since the future tracking environment is not known. Future problems with tracking at high instantaneous luminosity could increase the number of misconstructed tracks with very high IP significance and so a finite cut is used to minimise the effect of such potential problems on the SV-tagger.

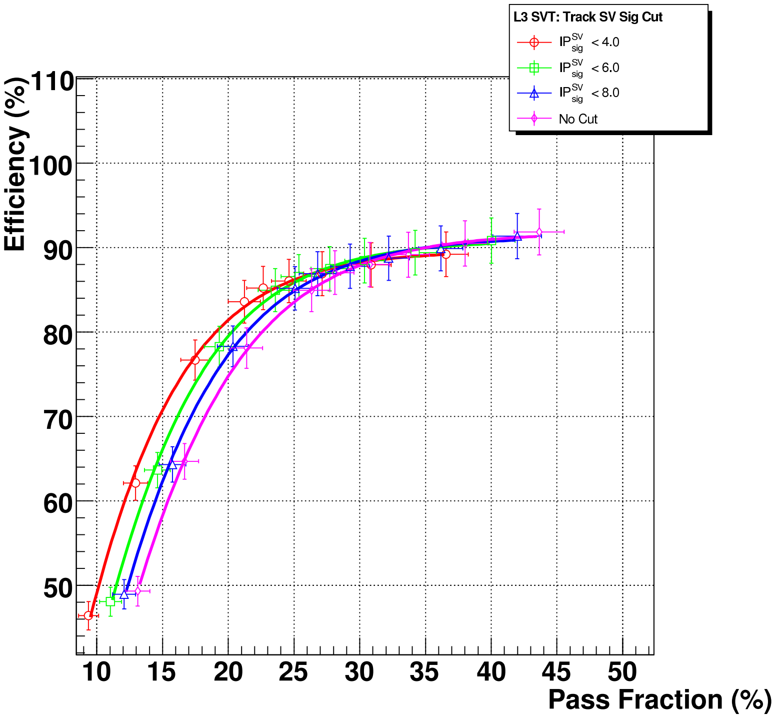

Figure 3.27: The performance of the SV-tagger using tracks with a maximum IP significance from the EP of 4 (red circles), 6 (green squares) and 8 (blue triangles), and with no cut (purple diamonds).

3.4.3.5 Maximum χ^2_{dof} Contribution

The secondary vertex candidates at each EP initially contain all tracks not attached to the PV. Tracks are removed one by one if they contribute more than χ^2_{max} to the vertex candidate χ^2_{dof}. This cut has the potential to remove mismeasured tracks but if too low it can also remove 'good' tracks and degrade the vertex reconstruction.

Cut values of 0.1, 0.5 and 1.0 were considered and the resulting performance is shown in Figure 3.28. The tool seems to be quite robust to variations in this cut, there is no significant difference in the tool performance over this range. A cut of 0.5 was chosen for the final configuration since it appears to have marginally better performance.

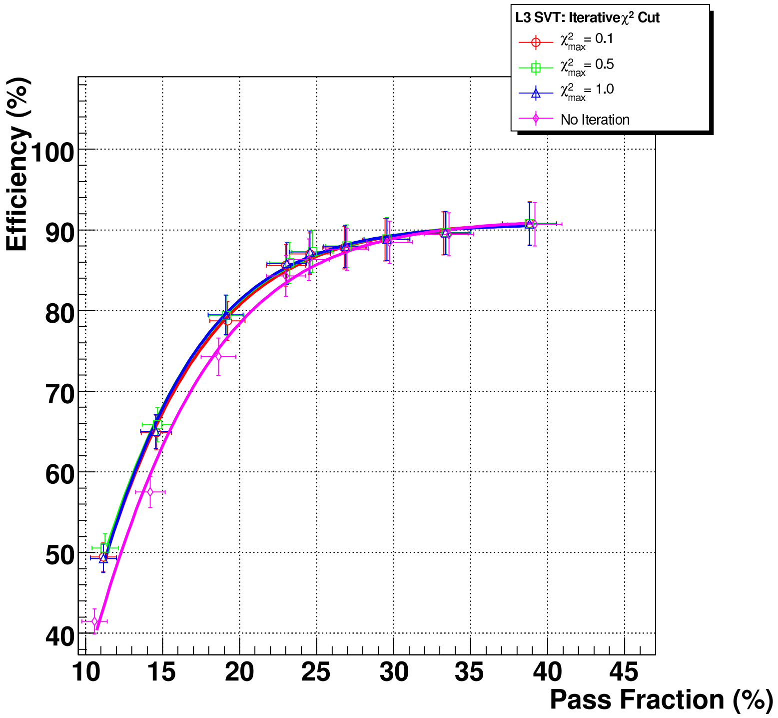

Figure 3.28: The performance of the SV-tagger as a function of the maximum χ^2_{dof} contribution of a track to the vertex candidate. Cut values of 0.1 (blue triangles), 0.5 (green squares) and 0.1 (red circles) are shown as well as the algorithm without this iterative track removal (purple diamonds) with statistical errors.

3.4.3.6 Vertex Candidate Separation

Vertices will be found at each EP which is a local minimum. Two local minima are merged if the χ^2_{dof} at each EP between the two minima is no more than χ^2_{split} above the χ^2_{dof} of the minima. When merged the minimum with the lowest χ^2 is taken as the remaining vertex candidate. The larger this cut the more clearly resolved adjacent vertex candidates are required to be. Shallow local minima will occur by chance in jets without real secondary vertices, this cut allows us to reject these. As we move along the jet axis the divergence of available tracks in the jet may mean minima become less well resolved and too large a threshold here may lead to a bias toward vertex candidates close to the PV and thus reduce efficiency.

Separation thresholds of 0, 0.5, 2 and 8 were considered, with the tool performance shown in Figure 3.29. The optimal performance was found when considering all local minima as vertex candidates, i.e. a separation threshold of 0.

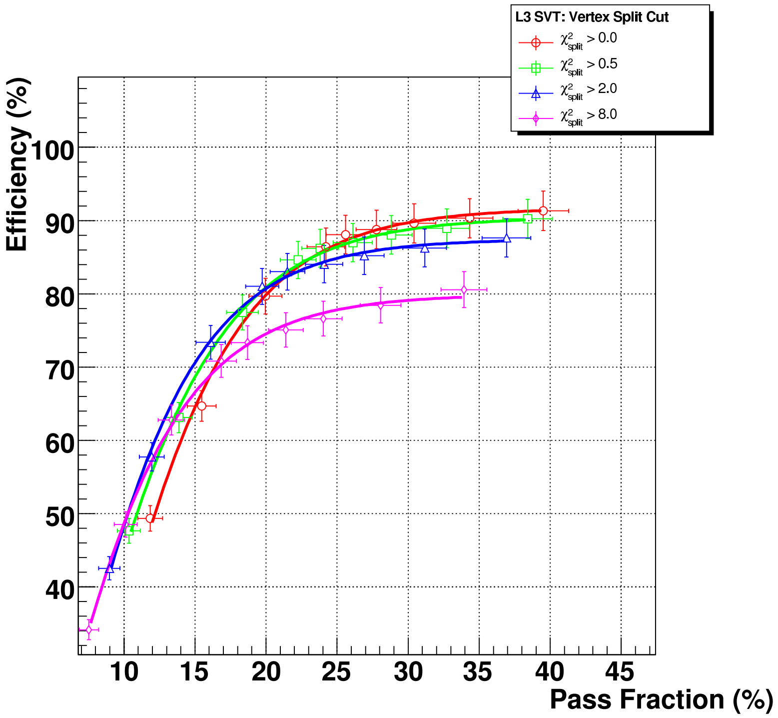

Figure 3.29: The SV-tagger performance with a range of vertex candidates separation cuts. Shown are thresholds of 0 (red circles), 0.5 (green squares), 2 (blue triangles) and 8 (purple diamonds) with statistical errors.

parameter	optimised value
track p_T	> 0.5 GeV
track χ^2	< 5.0
track PV significance	various working points
track SV significance	< 8.0
track χ^2 contribution	< 0.5
vertex candidate χ^2 separation	> 0.0

Table 3.3: The optimised parameter set for the SV-tagger.

3.4.4 Conclusion

The optimised parameters for the SV-tagger are summarised in Table 3.3. The significance of tracks from the PV, IP_{sig}^{PV}, gives a powerful handle on performance and is used to define several proposed operating points described in Table. 3.4. The performance of the optimised tool at the various working points is shown in Figure 3.30. At the $\sim 80\%$ efficiency performance benchmark the SV-tagger out-performs the IP-tagger, giving a rejection of 85% compared to the IP-tagger's $\sim 75\%$. The greater efficiency of the SV-tagger means that in the short term it might simply replace the IP-tagger. Ultimately combinations of the outputs of the taggers will be used to make maximum use of each tools ability to reject different background events. These combinations should also include the z-IP information and will probably make use of a multivariate technique such as the NN which has been so successful for offline b-tagging (as described in the following Chapter). The additional time required to run a second b-tagger is negligible since by far the largest amount of CPU time consumed for b-tagging is taken by running the tracking algorithm down to low p_T tracks and this need only be run once for both tools As the first step to including the SV-tagger output in the trigger decision, the tool has been added as a 'test trigger' to ensure that it behaves as expected when run online.

Name	IP^{SV}_{sig} cut	Efficiency (%)	Rate (%)
PV0	0	93	43
PV1	1	89	25
PV2	2	82	15
PV3	3	75	10
PV4	4	63	6
PV5	5	51	4

Table 3.4: The set of operating points with varying PV significance cuts with example efficiencies and rates for each. For a given efficiency the rate is estimated from the performance curve with an approximate error of ± 2 %.

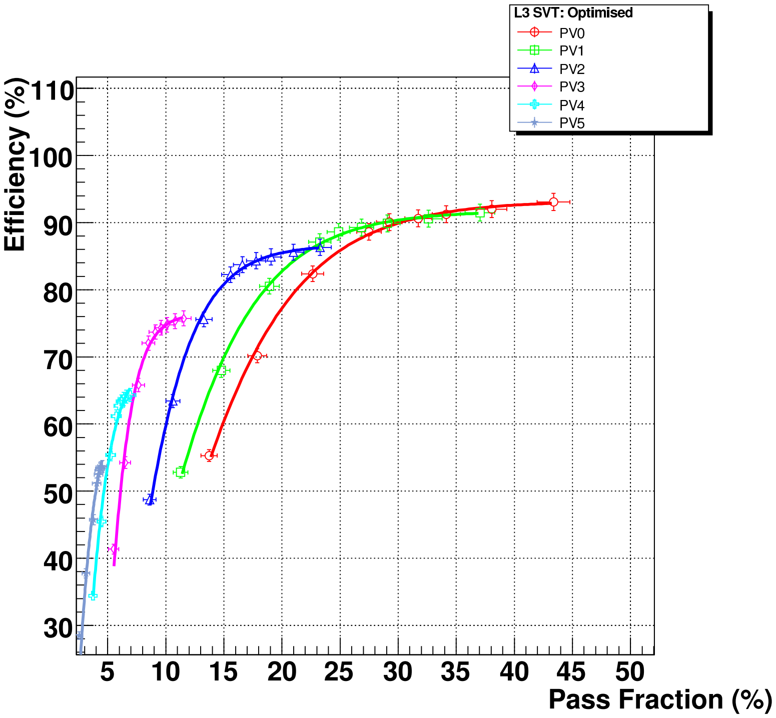

Figure 3.30: The performance of the SV-tagger at the six proposed operating points.

3.5 Conclusion and Outlook

The existing online *b*-tagger, the IP-tagger, has been retuned to take account of improved tracking in the latest trigger versions. The effect of modifications to the primary vertex finding and SMT unpacking tools on the IP-tagger has been studied, along with the effect of increasing instantaneous luminosity. A method for producing and processing a *b*-enhanced data sample for trigger studies has been developed and has been made available to the Level-3 studies group for future work. The use of the track z-information in the IP-tagger was investigated. An extension to the existing algorithm was developed, the z-track errors measured and the power of the tool investigated. Due to the larger stereo errors the improvement in efficiency for combined IP and z-IP tagging was limited. A new SV-tagger has been commissioned and is available for use in the next trigger version, this work is fully described in [59]. The SV-tagger offers a 10% improvement in absolute rejection over the IP-tagger, at 80% efficiency. The SV-tagger is running online as a ‘test trigger’. The combination of the SV-tagger with the existing IP-tagger is under consideration and will be explored in more detail in the future.

Chapter 4

Neural Network B-tagging

4.1 Introduction

As has been discussed in the preceding Chapter *b*-quarks are an important signal for several interesting rare processes, and highly effective *b*-tagging is therefore essential at all levels of data processing. Measurements of the top-quark and many Higgs searches rely on our ability to identify *b*-quarks.

The relatively long lifetime of the *b*-quark has been mentioned as providing an important handle for *b*-quark identification. The online lifetime-based algorithms described in the previous Chapter have offline analogues introduced in Chapter 2. The offline tools have less stringent restrictions in time and CPU consumption. The JLIP- and CSIP-taggers use the large impact parameter of tracks from the decay of a long lived particle (equivalent to the L3 IP-tagger). The secondary vertex tagger (SVT) tries to find the actual displaced decay vertices (like the Level-3 SV-tagger). The neural network (NN) tagger was developed [60, 61] to combine the existing lifetime taggers, using correlations between the individual tools to simultaneously increase efficiency and rejection.

Since the first round of NN development a major new version (p17) of the offline reconstruction algorithms has been released, the vertex finding tool was replaced and improvements were made to MC generation and simulation. As with the IP-tagger resolution refits, JLIP was retuned by its authors to take account of the new tracking environment. Following these updates to the NN-tagger's parent tools the NN was retrained and the output of the tool was certified for use in physics analyses. Certification is the process of measuring the performance of the tagger. A tagger must be certified before it can be used in a public DØ result. The certifiction

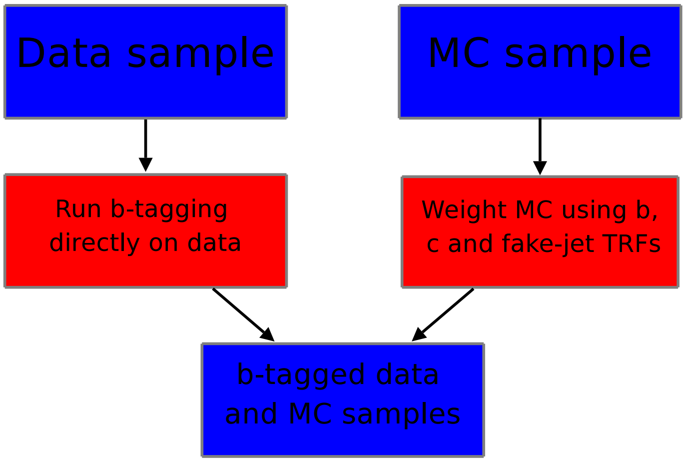

Figure 4.1: Schematic describing the correct application of b-tagging to data and MC.

must be documented and internally reviewed. The NN is trained using MC and the efficiency and fake rate are measured on data.

The work described in this Chapter consists of the retraining and certification of the latest version of the NN-tagger, which became the only certified b-tagging algorithm for use in all physics analyses at DØ. The remainder of Section 4.1 describes the certification procedure, some useful definitions and the data and MC samples used. Section 4.2 introduces the neural network and input variables, and describes the training of the algorithm. Section 4.3 describes the efficiency calculation using data and the assignment of errors. The fake rate and errors are calculated in Section 4.4. Sections 4.5 and 4.6 concern the overall performance and draw conclusions. Fuller details can be found in [62].

4.1.1 Introduction to the Certification Procedure

Because of the difficulty of simulating the tracking environment at a hadron collider the MC available at DØ does not perfectly reproduce the b-tagging performance in data. Tracking efficiency and quality are over estimated in MC with the result that the performance of the b-tagging algorithm on MC is unrealistically high. A system has been implemented to correct for this performance difference at the tagging level.

Figure 4.1 shows how b-tagging is performed, directly on data and using tag rate functions (TRFs) on MC. Four functions are derived to parametrise the complete data and MC performance on all jet flavours. These are expanded in terms of uncorrelated functions in p_T and η. The four functions are:

Scale Factor (SF) - The ratio of tag rates in data and MC. This can be used to correct the direct tagging rate in MC to the expected data rate. The SF is measured in the data and MC muonic b-jet samples only.

TRF$_b$ - The expected tag rate function (tagging efficiency) for inclusive b-jets in data. Calculated from inclusive b-jet MC and the scale factor.

TRF$_c$ - The expected tag rate function (tagging efficiency) for inclusive c-jets in data. Calculated from inclusive c-jet MC and the scale factor.

Fake Tag Rate (FTR) - The expected fake-jet tag rate (efficiency for light jets) in data. Calculated from the negative tag rate on data and correction factors from MC.

4.1.2 Physics Object Description

4.1.2.1 Taggability

Taggability is defined in Chapter 2. It is used to decouple detector performance from tagger performance, once a jet is taggable it is expected to have the same chance of being tagged as any other taggable jet with the same p_T and η, irrespective of its position in the detector.

4.1.2.2 Positive and Negative Tags

Positive tags are the standard outputs of the b-taggers used in analyses. Tracks and vertices are assigned a positive sign to indicate that they are consistent with a possible b-decay in the jet, a positive distance down the jet axis from the PV.

As with the Level-3 IP-tagger the individual offline b-taggers also produce negative tags which can be used to calculate the fake rate due to resolution effects. The spread of track IPs due to resolution will be symmetrical around the PV. Any tracks which cross the jet axis on the opposite side of the PV from the jet itself are inconsistent with having come from the jet. We can identify this sample of tracks and we call a tag assembled from such tracks a negative tag.

4.1.2.3 Away Tags

An away tag is a way of enhancing the b-content of a jet sample without making any cuts on the jet itself and therefore not biasing the jet sample in terms of p_T, η or probability to be tagged. Since b-quarks are pair produced tagging one jet produces a sample of events with an enhanced probability that another jet in the event contains a b-quark. For this Chapter an away tag is defined as a jet from a two jet event where the other jet has a JLIP probability less than 0.005.

Region	η range
Central (CC)	$0 < \lvert\eta\rvert < 1.2$
Inter-Cryostat Region (ICR)	$1.2 < \lvert\eta\rvert < 1.8$
End Cap (EC)	$1.8 < \lvert\eta\rvert < 2.4$

Table 4.1: calorimeter η regions

Skim	Skim criteria	Number of events	Number of jets after preselection
BID	One loose muon $p_T > 4$ GeV attached to a 0.7 cone jet	141M	15,277,180
EM	One electron with $p_T > 4$ GeV, missing $E_T < 10$ GeV	177M	48,055,113
QCD	Jet triggers, not in EM skim	72M	106,417,629
COMB	Combination of EM and QCD skims	249M	154,472,742

Table 4.2: The data skims with the selection criteria, number of events and number of usable jets.

4.1.2.4 Calorimeter Regions

The calorimeter, described in Section 2.2.3, consists of three regions: the central calorimeter (CC), the end cap (EC) and the inter-cryostat region (ICR). Differences in calorimetry and tracking response can make it useful to break down studies into these three regions. The regions will be used later in this Chapter and are defined in terms of η in Table 4.1.

4.1.3 Data and MC Samples

The data sample used for the certification was collected between July 2002 and February 2006. Any runs or parts of runs with problems in the readout of any detector components were excluded. Three data 'skims' were used, where a skim is a data set defined by some simple criteria. The 'BID' skim had an enhanced b-content and was used for the efficiency measurement. The 'EM' and 'QCD' skims, and the combination of the two, were used for the fake rate studies. The number of events and skimming criteria are described in Table 4.2.

Table 4.3 outlines the MC samples used. The QCD and Z^0 samples were produced using Pythia [58] version 'v6.319'. The common $t\bar{t}$ samples produced using

Sample	No. Events
inclusive $t\overline{t}$	571,750
$Z^0 \to b\overline{b}$	105,250
$Z^0 \to b\overline{b}$ with μ	105,750
$Z^0 \to c\overline{c}$	107,250
$Z^0 \to c\overline{c}$ with μ	107,750
$Z^0 \to q\overline{q}$ (q = u,d,s)	103,750
$Z^0 \to q\overline{q}$ with μ (q = u,d,s)	107,000
QCD $q\overline{q}$	972,500
QCD $b\overline{b}$	265,000
QCD $c\overline{c}$	239,287

Table 4.3: MC samples used for certification studies.

Pythia were combined with samples produced by the top physics working group using Alpgen [63]. MC jets are called *b*-jets if they have a *b*-hadron within $\Delta R < 0.5$. MC *c*-jets are defined similarly and the remainder are 'light' (uds and gluon) jets. Jets are identified as 'muonic' if a muon, as defined in Section 2.4.1, is attached to a track within $\Delta R < 0.5$ of the jet axis. QCD MC samples generated with different p_Ts are merged into a continuous sample weighted to produce a smoothly falling jet p_T spectrum, similar to that seen in data. The combined samples are formed by simply combining all available jets irrespective of the MC process used to produce them into the largest possible generic sample of each type of jet. Table 4.4 shows the number of jets passing the jet preselection in the individual, combined and merged flavour samples.

4.1.4 Operating Points

In the following sections the NN performance, the fake-rate, b/c tag-rate, MC SF and associated systematic and statistical errors are evaluated at each of twelve operating points. Table 4.5 gives the NN output cut value for each operating point. The values are chosen to cover a broad range of efficiency and purity, broader than previously available with any other tagger. The plots in the following sections refer to the 'Tight' operating point, those for the other eleven points were also produced but are not shown. The results of these studies for all operating points are included in the tables and in the performance curves. The full set of plots for the remaining operating points can be found in [62].

Flavour	Sample	Number of jets
b	$Z \to b\overline{b}$	130,220
	$t\overline{t} \to b$	1,662,929
	QCD $b\overline{b}$ (Merged)	90,252
	b (Combined)	2,349,755
$b \to \mu$	$Z \to b\overline{b} \to \mu$	42,726
	$t\overline{t} \to b \to \mu$	175,579
	$b \to \mu$ (Combined)	273,281
c	$Z \to c\overline{c}$	137,740
	$t\overline{t} \to c$	404,753
	QCD $c\overline{c}$ (Merged)	90,441
	c (Combined)	1,076,054
$c \to \mu$	$Z \to c\overline{c} \to \mu$	47,986
	$t\overline{t} \to c \to \mu$	17,715
	$c \to \mu$ (Combined)	91,824
uds	$Z \to q\overline{q}$	152,590
	QCD Fake (Merged)	239,118
All	QCD All	2,086,603

Table 4.4: Number of jets of each flavour in each MC sample after processing and jet selection. The merged samples are produced by a weighted combination of samples generated at different p_Ts in order to have a smoothly falling distribution. The combined samples are simply the total sample of all jets of each flavour, equally weighted regardless of origin.

L6	L5	L4	L3	L2	Loose
0.1	0.15	0.20	0.25	0.325	0.45

oldLoose	Medium	Tight	VeryTight	ExtraTight	MegaTight
0.5	0.65	0.775	0.85	0.9	0.925

Table 4.5: The twelve NN operating point names and the minimum NN output required to pass the tagger at each operating point.

4.2 Building the Network

4.2.1 Introduction to Neural Networks

A neural network is a mathematical construction to produce a discriminating variable from a multi-variable input [64]. A multi-layer perceptron is a type of feed-forward NN consisting of processing nodes in layers which relay data in one direction only: from inputs to output. A ROOT package, TMultilayerPerceptron [65], was used to train and test the neural network. Subsequently a hard-coded C++ version of the network was produced for use in the reconstruction code. The multilayer perceptron is composed of an input layer, one or more hidden layers and an output layer. For the binary discrimination required here only a single output node is required [66]. The original neural networks were developed in analogy to organic neural networks, like the brain, as such the nodes are known as neurons and the links as synapses. Each neuron is attached to all nodes in adjacent layers by synapses with weights w_{ij}, where i and j specify the neurons being linked. An additional bias, w_{0j}, is associated with each neuron j. The neuron j combines the outputs from the nodes in the preceding layer, y_i, in a linear sum to give x_j:

$$x_j = w_0 j + \sum_i w_{ij} y_i \tag{4.1}$$

The neuron then outputs either this linear sum or a function of the sum. The choice of function varies, here a sigmoid is used. The output, z_j, of the neuron j is then:

$$z_j = x_j \quad \text{(for output neurons)} \tag{4.2}$$

or

$$z_j = \tfrac{1}{1+e^{x_j}} \quad \text{(for other neurons)} \tag{4.3}$$

Neural networks are also distinguished by their ability to be progressively trained to discriminate between samples. Training can reveal correlations between variables which are not immediately visible and can help to gain the most discrimination from a set of variables. Initially the weights for each synapse and neuron are set randomly between -0.5 and 0.5. The training sample consisting of equal numbers of signal and background events, or unequal numbers weighted to be equal, is processed and the outcomes compared to the desired outcomes. The error is proportional to the number of incorrect responses. The total training error, E, is given by:

$$E = \sum_p \frac{1}{2} w_p (o_p - t_p)^2 \tag{4.4}$$

where o_p is the actual output, t_p is the desired output and w_p is the weight attached to the event p. When training the first derivative of this error with respect to the synapse and neuron weights is calculated and the weights are adjusted with the aim of minimising the number of incorrect outcomes. Various methods of reaching the minimum are possible, and each should eventually produce equivalent results. Each time the error is calculated and the weights are adjusted a training 'epoch' is said to have passed. The Broyden, Fletcher, Goldfarb, Shanno (BFGS) [67] algorithm was seen to require the lowest number of epochs to reach the best configuration during the original NN-tagger development [61]. The BFGS method relies on the theory of unconstrained minimisation [68]. In a multidimensional 'space' defined by the weights, $\vec{w}$, a direction, $\vec{s}$, is found from the derivative of E with respect to each weight, ∇E. Rather than taking the steepest gradient the BFGS method chooses a direction by a multidimensional version of the Newton-Raphson method [69]. The weights for the $t+1$ epoch are given by:

$$\vec{w_{t+1}} = \vec{w_t} + \alpha \vec{s_t} \tag{4.5}$$

where α is chosen such that $E(\vec{w_t} + \alpha \vec{s_t})$ is minimised. This process is repeated for each training epoch. A second set of events is used to determine a second error calculated identically to the training error, this quantity is called the test error. The test error is independent of the training error used to improve the weights and is used to gauge the point at which the training is sufficient, to prevent over training.

4.2.2 The NN-tagger Input Variables

The variables to be used in the NN were selected for their discrimination power during the first round of NN certification [61]. This design for the NN structure was retained along with the optimised number of hidden layers; twice as many as input variables.

The variables input to the NN are:

SVT DLS - The decay length significance of the secondary vertex with the largest decay length significance with respect to the PV (known as the 'most significant' SV).

CSIP comb - A weighted combination of the number of tracks in the various CSIP categories. Since neural networks prefer continuous variables it was necessary to construct CSIP comb from the small discreet values of the CSIP variables, as stated in Equation 4.6.

$$CSIPcomb = 6 \times 3s + 4 \times 2s + 3 \times 3w + 2 \times 2w \tag{4.6}$$

where:

3s - The number of tracks with a positive impact parameter significance greater than 3.

2s - The number of tracks with a positive impact parameter significance greater than 2.

3w - The number of tracks with a negative impact parameter significance greater than 3 within a 1.15 radian cone around the jet axis.

2w - The number of tracks with a negative impact parameter significance greater than 2 within a 1.15 radian cone around the jet axis.

JLIP prob - The probability that the jet originated from the PV, based on the IP significance of tracks in the jet. Notably this variable is one for background and zero for signal, the opposite of the NN output.

SVT χ^2_{dof} - The χ^2 per degree of freedom of the most significant secondary vertex fit.

SVT N_{tracks} - The number of tracks used in the most significant secondary vertex.

SVT Mass - An estimated mass of the most significant secondary vertex. The mass is calculated from the combined 4-momenta of all the tracks assuming that all particles are pions.

SVT Num - The number of secondary vertices reconstructed in the jet.

If any of the tools do not return a tag for an event the input variable is set to zero or, in the case of the χ^2_{dof}, to 75. Figure 4.2 shows the distributions of the variables in the QCD data skim and in QCD light-jet and direct $b\bar{b}$ MC.

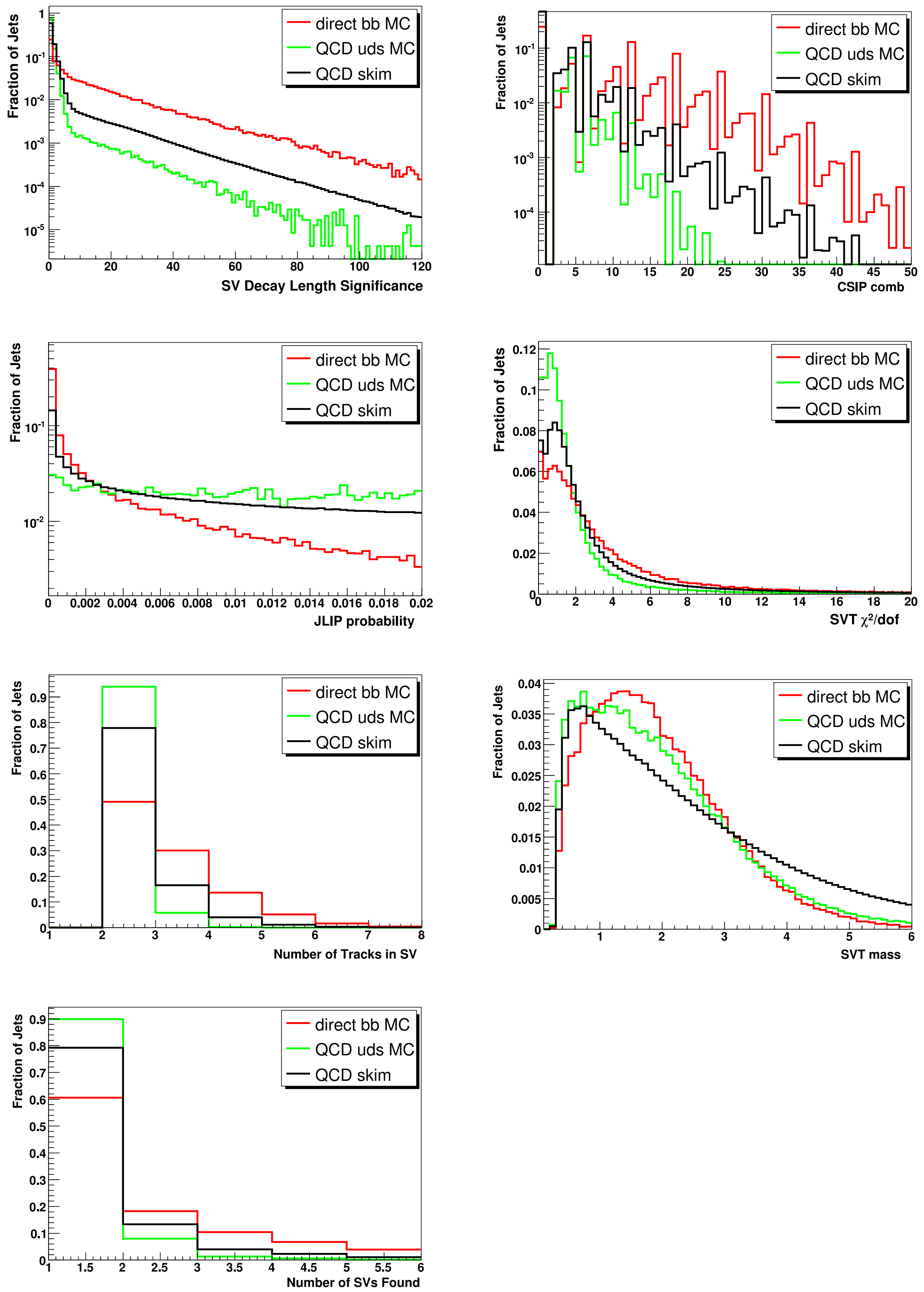

Figure 4.2: The NN input variables for MC b-jets, MC QCD light jets and for the QCD data skim. Each distribution is normalised to unit area for comparison.

4.2.3 Negative Tag Input Variables

The lifetime taggers all produce a negative tag (NT) value, using tracks and vertices which are not compatible with the evolution of a real b-jet. The negative tags are defined differently for each algorithm:

CSIP NT - Similar to CSIP comb, but with the sign of the IP conditions reversed.

JLIP NT - Identical to JLIP probability except that negative signed IP tracks are used instead of positive ones.

SVT NT - The SV variables are produced as with the positive tag but from vertices with a negative decay length within $\Delta R < 0.5$ of the jet axis.

4.2.4 Tagger Preselection

Only jets which pass a relatively loose logical-OR of the individual taggers are considered by the NN tagger. A jet is required to have either a SVT decay length significance (DLS) > 2.5, a JLIP probability < 0.02 or CSIP comb > 8. Jets which fail the preselection fail the NN tagger and are assigned a NN output of -1. The preselection must be loose since any efficiency lost at this stage cannot be recovered. However to reserve the power of the NN to reject the most b-like fakes, the preselection is as tight as it can be without significantly damaging the efficiency. With such a preselection the NN is being trained to reject those fakes which otherwise appear to be a real tag to at least one of the individual taggers. The preselection was optimised when the original NN was developed [61] and was not re-optimised for this round of certification.

4.2.5 Neural Network Training

The network was trained using QCD light jet and $b\bar{b}$ MC samples weighted to equal numbers after the preselection. The signal MC contained 270,000 jets with 220,000 passing the preselection and the background contained 470,000 with 40,000 passing the preselection. The signal and background samples were divided into two halves to give independent samples for training and testing the network. The number of training epochs was varied between 50 and 2000; 400 epochs was seen to be sufficient. The network weighting with the fewest false results on the test samples was taken to be the final best network. The number of incorrect results as a function of training epoch for the two independent samples is shown in Figure 4.3.

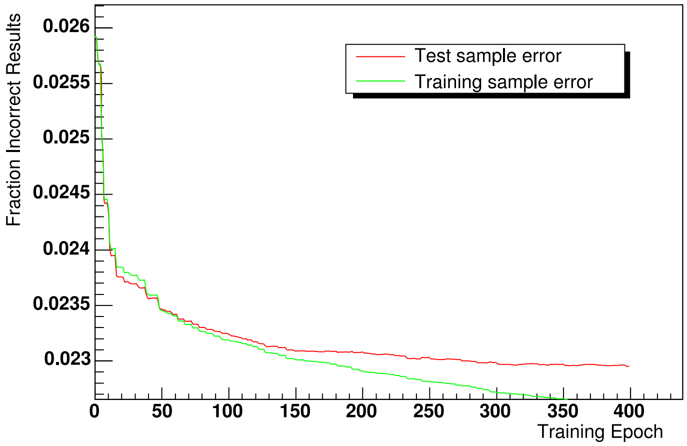

Figure 4.3: The fraction of incorrect results for the training and test samples as a function of training epoch. Four hundred epochs is seen to be sufficient to reach the minimum error on the test sample.

4.2.6 MC Performance

The output of the trained NN tagging algorithm when run on MC $b\bar{b}$ and QCD light jets is shown in Figure 4.4. Only those events which pass the preselection appear in this plot, that is those events which look like a plausible b-jet to at least one of the existing taggers. The separation between signal and background should be interpreted in this light.

4.3 Data Certification for b-Efficiency

The complete method to derive inclusive b and c-jet TRFs is shown in Figure 4.5. The data/MC scale factor (SF) is calculated from muonic b-jets in data and MC. The SF is then used to correct the inclusive b- and c-jet efficiencies measured in MC to give the inclusive data tag rate functions (TRFs). The first stage, the derivation of a TRF for muonic b-jets in data, requires the use of the system-8 method [70].

4.3.1 The System-8 Method

The system-8 method (S8) can be used to measure the b-efficiency of a data sample. The method employs two taggers and two data samples to form a set of eight simultaneous equations. The solution to the equations gives the b-content of the samples and the efficiencies of the taggers. Since the equations are symmetrical they can only be solved if the taggers and samples are sufficiently different from each other. The taggers should be uncorrelated with different b-efficiencies and the samples must have different b-content.

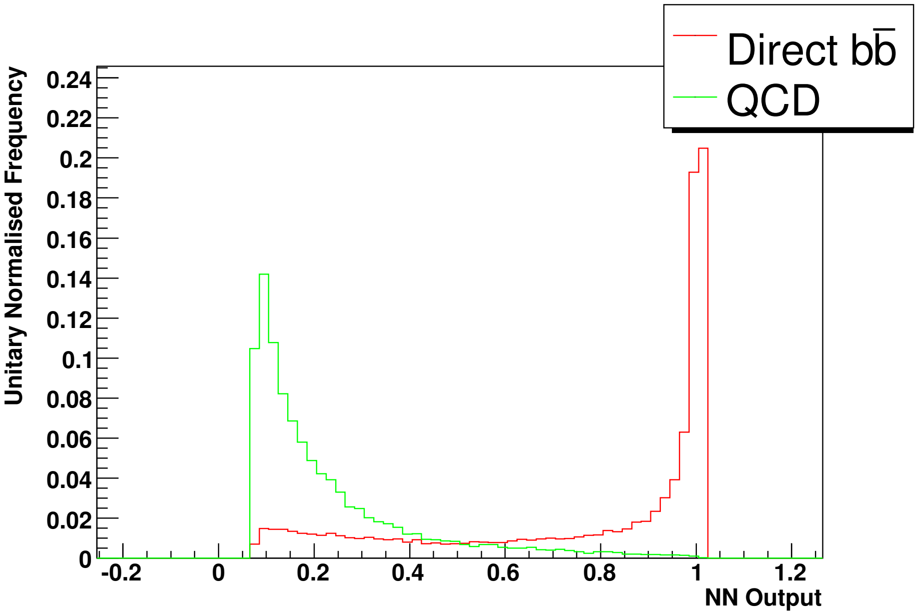

Figure 4.4: Output of the NN on direct $b\bar{b}$ and QCD MC.

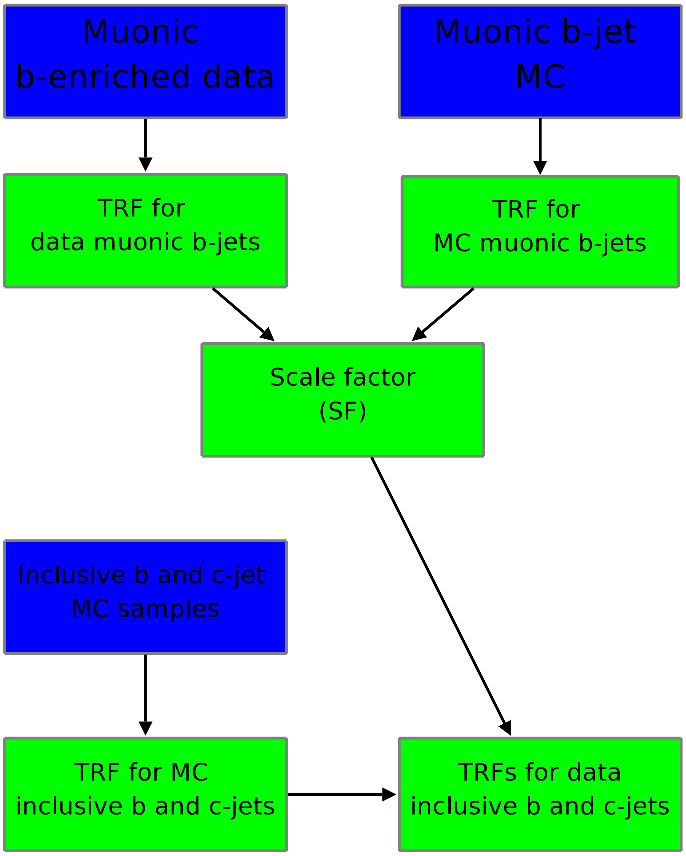

Figure 4.5: Schematic describing the process of measuring the efficiency, the TRFs, for inclusive b- and c-jets.

The two taggers used are the NN and soft lepton tagger (SLT). The SLT operating point is chosen such that the SLT efficiency is different from the NN efficiency at all operating points for all p_T and η bins. A jet is tagged by the SLT if it has an associated muon with $p_{T_{Rel}}$ greater than 0.5 GeV. The efficiency of the SLT with this cut was measured to be $\sim 80\%$ on the muonic $b\bar{b}$ MC sample.

The first data sample is the BID skim (see Table 4.2). This skim requires 4 GeV muons which must be matched to a jet above 15 GeV. The second sample consists of jets with an away tag.

To account for remaining correlations between the taggers and differences between the samples, four correlation coefficients are introduced each of which we expect to be close to unity. The values of these quantities are measured on MC and the errors on these measurements contribute to the overall systematic uncertainty on the S8 method.

The eight simultaneous equations are:

$$n = n_b + n_{udsc} \tag{4.7}$$

$$p = p_b + p_{udsc} \tag{4.8}$$

$$n^{SLT} = \epsilon_b^{SLT} n_b + \epsilon_{udsc}^{SLT} n_{udsc} \tag{4.9}$$

$$p^{SLT} = \epsilon_b^{SLT} p_b + \epsilon_{udsc}^{SLT} p_{udsc} \tag{4.10}$$

$$n^{NN} = \epsilon_b^{NN} n_b + \epsilon_{udsc}^{NN} n_{udsc} \tag{4.11}$$

$$p^{NN} = \beta\epsilon_b^{NN} p_b + \alpha\epsilon_{udsc}^{NN} p_{udsc} \tag{4.12}$$

$$n^{SLT,NN} = \kappa_b \epsilon_b^{SLT} \epsilon_b^{NN} n_b + \kappa_{udsc} \epsilon_{udsc}^{SLT} \epsilon_{udsc}^{NN} n_{udsc} \tag{4.13}$$

$$p^{SLT,NN} = \kappa_b \beta \epsilon_b^{SLT} \epsilon_b^{NN} p_b + \kappa_{udsc} \alpha \epsilon_{udsc}^{SLT} \epsilon_{udsc}^{NN} p_{udsc} \tag{4.14}$$

where n and p are the numbers of jets in the muonic and b-enhanced samples respectively. The ϵ are the tagging efficiencies. The superscripts refer to the taggers each sample has passed and the subscripts to the flavours, b or light.

The remaining coefficients describe the correlations:

- α - The ratio of the NN *udsc*-tagging efficiency for the two samples
- β - The ratio of the NN b-tagging efficiencies for the two samples
- κ_b - To account for the correlation between the NN tagger and SLT efficiencies on b-jets.
- κ_{udsc} - To account for the correlation between the NN tagger and SLT efficiencies on light jets.

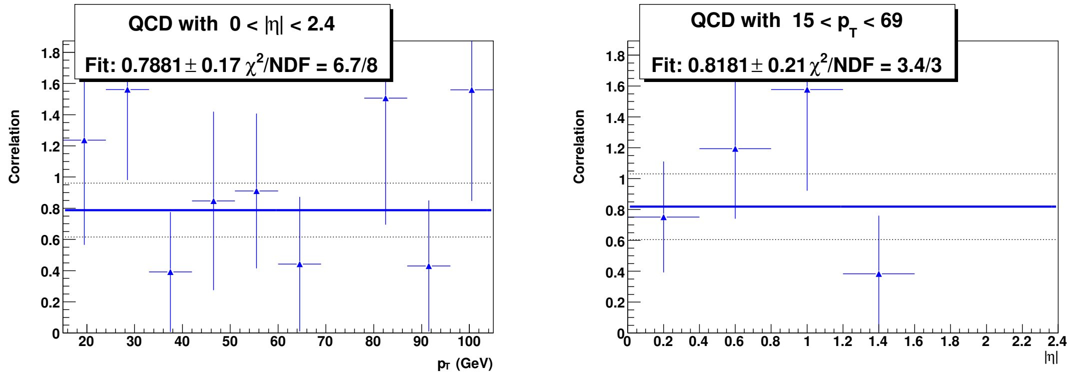

Figure 4.6: The ratio of tag rates in the all-*udsg* and away tagged *udsg* samples and the fit to the ratio which gives the parameter α. The errors on the fit are shown as dotted lines. The actual tag rates for the two samples are very small and are not shown. The large errors reflect the limited size of the tagged udsg-jet sample.

Since it is necessary to accurately know the jet flavours in order to calculate these correlation coefficients, they must be estimated from MC. In addition the effect of changing the $p_{T_{Rel}}$ cut in the SLT definition is studied. This factor is not included in the equations given above but the variation is used in the error calculation.

4.3.2 Measuring the Correlation Coefficients

4.3.2.1 α

The ratio of the tagging efficiencies for light, uds and gluon, jets in the two samples is parametrised in the S8 equations by α. The efficiencies are expected to be similar, in a given p_T and η bin, and therefore α should be close to one. The NN tag rates for light jets are calculated using the QCD light flavour MC for the entire sample and for the subsample of jets with an away tag. The ratio α is shown in Figure 4.6 along with the light-jet efficiencies in p_T and η bins. The statistical error on the α calculation takes into account the overlap between the two samples. The systematic error due to α on the b-efficiency from S8 was estimated by varying α within its statistical error.

4.3.2.2 β

The parameter β is the ratio of the b-efficiencies in the muonic b-jet sample and the away tagged muonic b-jet sample. The ratio β calculated using the combined $b \to \mu$ MC sample is shown in Figure 4.7, the two individual efficiencies, in the inclusive and away tagged sub-samples, are also shown. As with α the statistical error accounts for the correlation between events shared by the two samples. The

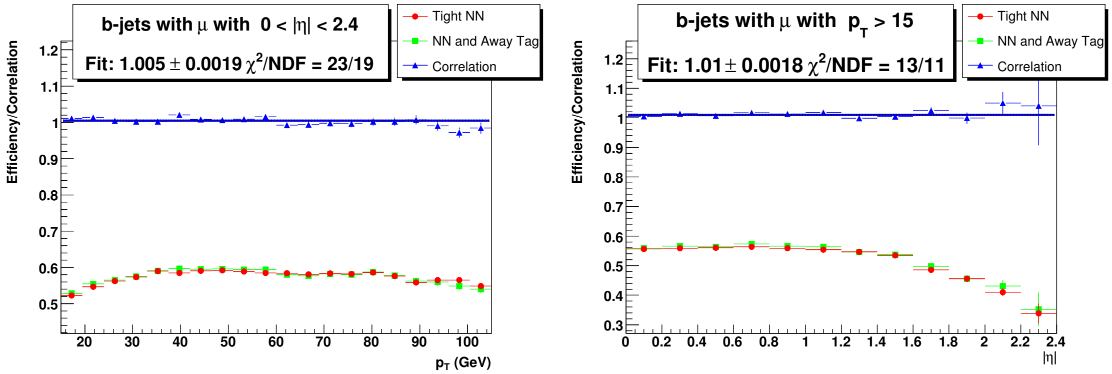

Figure 4.7: The NN Tight tag rates in the inclusive and away tagged muonic b-jet samples and the ratio β.

sample dependence of the b-efficiencies and hence β provides an additional source of error. The efficiency ratio β was recalculated with the $Z^0 \rightarrow b\bar{b}$ and $t\bar{t}$ MC samples. The sample dependence error was taken to be half the deviation from unity of the ratio of these β values parametrised in p_T and η. The statistical and sample dependence errors were added in quadrature to give the combined error. The systematic on the b-efficiency due to uncertainty in β was then calculated by resolving the S8 equations while varying β within this combined error.

4.3.2.3 $\kappa_{b,udsgc}$

The correlations between the two taggers for the light and b-efficiencies are parametrised as κ_{udsgc} and κ_b. The κ factors are calculated by comparing the efficiency to tag an event with both taggers to the product of the individual tagging efficiencies, as shown in Equation. 4.15:

$$\kappa_{b,udsgc} = \frac{\epsilon_{b,udsgc}^{NN+SLT}}{\epsilon_{b,udsgc}^{NN} \times \epsilon_{b,udsgc}^{SLT}} \tag{4.15}$$

Figure 4.8(top) shows κ_b as measured in the combined muonic b-jet MC sample for p_T and η. A linear fit is performed for the p_T parametrisation and the fit error is determined from the covariance matrix. The resulting value for κ_b is constant in η and has a small p_T dependence. As a result the fit error and subsequently the systematic due to κ_b are each a function of p_T.

The error from sample dependency of κ_b is calculated by comparison of κ_b derived from the $Z \rightarrow b\bar{b}$ and $t\bar{t}$ samples. The uncertainty is taken to be half the average deviation from unity of the ratio of κ_b values derived from these two samples in the p_T and η projections. Due to the different p_T composition of the samples and

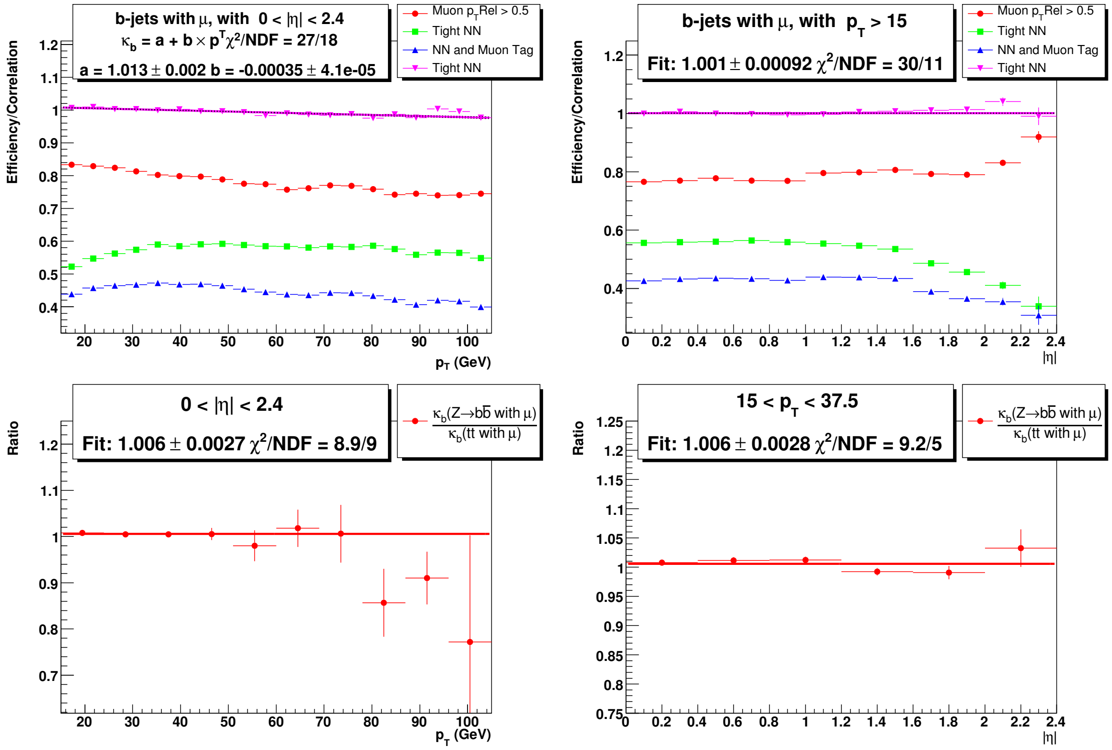

Figure 4.8: The tag rates of the muon, NN Tight and combined tags for the muonic b sample and the ratio κ_b are shown (top). The systematic error on κ_b derived from the ratio of κ_b calculated with b-jets with muons and $t\bar{t}$ with muons.

the p_T dependence of κ_b, the p_T range of the η projection is limited to between 15 and 37.5 GeV. Figure 4.8(bottom) shows the ratio of κ_b values used for the sample dependence error. The total error due to κ_b is the combination in quadrature of the sample dependence and statistical errors.

κ_{udsgc} is measured with a similar method using the combined muonic c-jet samples. Again a sample dependence error is calculated using the individual muonic Z and $t\bar{t}$ samples containing c-quarks. The sample dependence ratio can be seen in Figure 4.9. The sample dependence error was estimated as half the average deviation of the ratio from unity and was added in quadrature to the statistical error to give the total error due to κ_{udsgc}.

4.3.2.4 Variation with SLT Operating Point

The S8 equations refer to a single operating point for both taggers. Since the choice of SLT operating point is somewhat arbitrary (provided it is efficient enough) it is necessary to quantify the effect of this choice. The NN tagger efficiency was recalculated with a range of SLT operating points, varying the minimum $p_{T_{Rel}}$ requirement

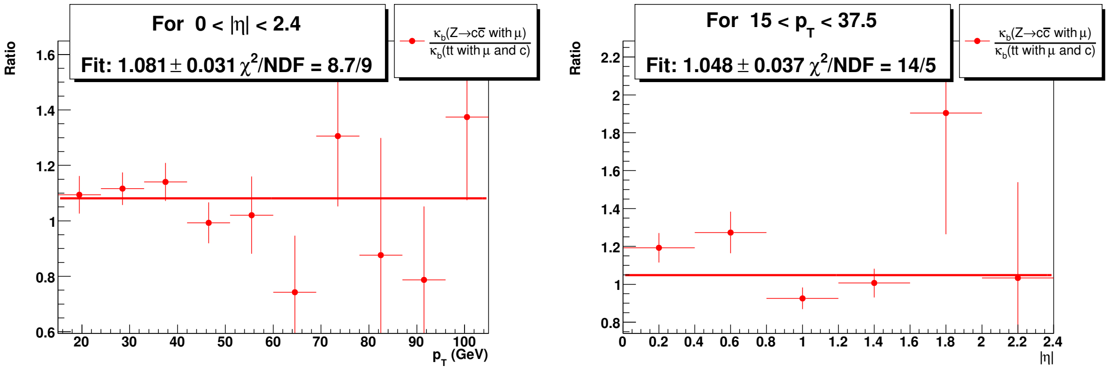

Figure 4.9: The systematic error on κ_{udsgc} derived from the ratio of κ_{udsgc} calculated with $Z \rightarrow c\overline{c}$ with muons and $t\overline{t}$ with c-quarks and muons.

between 0.3 and 0.7 GeV. The κ correlations were recalculated for each SLT operating point and the overall effect on the efficiency was taken as a further systematic error.

4.3.2.5 Correlation Coefficient Values

The values and errors from the p_T and η parametrisations of the coefficients are given in Tables 4.6 and 4.7 for the twelve NN operating points. Since κ_b is p_T dependent the values quoted here are taken at the average sample p_T. When the S8 equations are solved to give efficiency parametrisations in p_T and η bins the average p_T for the given bin is used to calculate the correlations. Systematic errors on the b-efficiency resulting from varying these coefficients within their errors are calculated in the following section.

Coefficients	**L6**	**L5**	**L4**	**L3**
α	0.967 ± 0.096	0.937 ± 0.102	0.957 ± 0.116	0.918 ± 0.124
β	1.010 ± 0.001	1.010 ± 0.001	1.010 ± 0.001	1.010 ± 0.002
κ_b	0.999 ± 0.002	0.999 ± 0.001	1.000 ± 0.001	1.000 ± 0.001
κ_{udsgc}	1.000 ± 0.018	0.998 ± 0.022	0.999 ± 0.026	1.000 ± 0.025

Coefficients	**L2**	**Loose**	**oldLoose**	**Medium**
α	0.766 ± 0.127	0.819 ± 0.152	0.840 ± 0.160	0.791 ± 0.183
β	1.010 ± 0.001	1.010 ± 0.001	1.010 ± 0.002	1.010 ± 0.002
κ_b	1.000 ± 0.002	1.000 ± 0.002	1.000 ± 0.003	1.000 ± 0.003
κ_{udsgc}	1.000 ± 0.022	1.000 ± 0.021	0.999 ± 0.025	0.995 ± 0.027

Coefficients	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
α	0.816 ± 0.213	0.904 ± 0.245	1.070 ± 0.326	0.951 ± 0.338
β	1.010 ± 0.005	1.010 ± 0.006	1.010 ± 0.007	1.010 ± 0.008
κ_b	1.000 ± 0.003	1.000 ± 0.002	1.000 ± 0.002	1.000 ± 0.001
κ_{udsgc}	0.987 ± 0.033	0.979 ± 0.025	0.958 ± 0.026	0.942 ± 0.029

Table 4.6: The S8 correlation coefficients with associated errors as derived in the MC samples from the η parametrisation.

Coefficients	**L6**	**L5**	**L4**	**L3**
α	1.060 ± 0.079	1.010 ± 0.086	0.921 ± 0.091	0.897 ± 0.098
β	1.000 ± 0.001	1.000 ± 0.001	1.010 ± 0.001	1.000 ± 0.002
κ_b	1.000 ± 0.002	1.000 ± 0.001	1.000 ± 0.001	1.000 ± 0.001
κ_{udsgc}	1.000 ± 0.018	1.000 ± 0.022	1.000 ± 0.027	1.010 ± 0.025

Coefficients	**L2**	**Loose**	**oldLoose**	**Medium**
α	0.786 ± 0.104	0.864 ± 0.126	0.789 ± 0.125	0.829 ± 0.156
β	1.010 ± 0.002	1.000 ± 0.002	1.000 ± 0.002	1.000 ± 0.002
κ_b	1.000 ± 0.002	1.000 ± 0.002	1.000 ± 0.003	1.000 ± 0.003
κ_{udsgc}	1.010 ± 0.022	1.010 ± 0.021	1.000 ± 0.025	0.998 ± 0.027

Coefficients	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
α	0.796 ± 0.174	0.940 ± 0.205	0.877 ± 0.224	0.904 ± 0.249
β	1.000 ± 0.005	1.000 ± 0.006	1.010 ± 0.007	1.010 ± 0.008
κ_b	1.000 ± 0.003	0.999 ± 0.002	0.999 ± 0.002	0.999 ± 0.001
κ_{udsgc}	0.989 ± 0.033	0.981 ± 0.025	0.960 ± 0.026	0.943 ± 0.030

Table 4.7: The S8 correlation coefficients with associated errors as derived in the MC samples from the p_T parametrisation.

4.3.3 The S8 b-efficiency

The S8 equations were solved numerically, using TMinuit [71], for the whole sample and in bins of p_T and η to generate two one dimensional profiles of tagging efficiency. The systematic errors due to uncertainty in the S8 correlation coefficients were estimated by varying the coefficients within their errors. The error estimation was performed on both the complete sample and on each of the p_T and η bins. Figure 4.10 shows the systematic error due to each of the correlation coefficients for each p_T and η bin for the Tight NN operating point. The total systematic error from the correlation coefficients is also given. A constant fit is made to the total error over p_T and η and the fit values are used to generate the $\pm 1\sigma$ curves used to calculate the statistical error in Section 4.3.4.3. The systematics have little η dependence and (at low p_Ts) tend to reduce slightly with increasing p_T. Using the systematic calculated on the whole sample neglects this p_T correlation and would underestimate the total systematic error.

Table 4.8 shows the overall efficiencies (integrated over p_T and η) and the systematic errors contributed by the correlation coefficients over the whole sample for each of the NN operating points. The sum in quadrature of these errors gives the full sample total systematic error which ranges between ~ 1 and 2% for b-efficiencies of 37 to 77%.

4.3.3.1 Efficiency Parametrisation

The efficiencies were parametrised in p_T and η using the functions given in equations 4.16 and 4.17.

$$\epsilon\left(p_T\right) = \frac{c}{1 + ae^{-bp_T}} \tag{4.16}$$

$$\epsilon\left(\eta\right) = d + e\eta + f\eta^2 + g\eta^3 + h\eta^4 \tag{4.17}$$

where a,b,c...h are parameters to be determined by the fitting proceedure. It is assumed in the standard DØ certification procedure that the p_T and η dependencies can be factored out, i.e. any correlation is minimal. Previous studies with JLIP [42] and an earlier version of the NN [72] have shown this to be a good assumption. A 2-D parametrisation was thus formed by the product of the efficiencies:

$$\epsilon\left(p_T, \eta\right) = \frac{1}{\epsilon^{all}} \times \left(\frac{c}{1 + ae^{-bp_T}}\right) \times \left(d + e\eta + f\eta^2 + g\eta^3 + h\eta^4\right) \tag{4.18}$$

where ϵ^{all}, the overall efficiency, is used to give the correct normalisation.

Error	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
Efficiency	77.0%	74.9%	72.2%	69.6%	65.9%	60.8%
Total syst.	1.7 %	1.6 %	1.5 %	1.3 %	1.3 %	1.4 %
α	0.04%	0.02%	0.02%	0.03%	0.02%	0.03%
β	0.14%	0.14%	0.17%	0.21%	0.18%	0.18%
κ_b	0.66%	0.51%	0.56%	0.48%	0.73%	0.98%
κ_{cl}	0.66%	0.66%	0.59%	0.46%	0.29%	0.25%
$p_{T_{Rel}}$	1.38%	1.33%	1.21%	1.11%	1.01%	0.91%
Error	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
Efficiency	59.3%	53.7%	47.6%	43.3%	39.5%	37.1%
Total syst.	1.5 %	1.3 %	1.5 %	1.5 %	1.3 %	1.4 %
α	0.02%	0.00%	0.02%	0.04%	0.07%	0.11%
β	0.20%	0.26%	0.55%	0.78%	0.80%	0.96%
κ_b	1.23%	1.21%	1.21%	0.96%	0.70%	0.60%
κ_{cl}	0.25%	0.23%	0.22%	0.18%	0.16%	0.16%
$p_{T_{Rel}}$	0.69%	0.46%	0.70%	0.85%	0.78%	0.86%

Table 4.8: Relative systematic uncertainties on the NN tag efficiency determined with the S8 method. The total systematic uncertainty was determined by adding the individual uncertainties in quadrature.

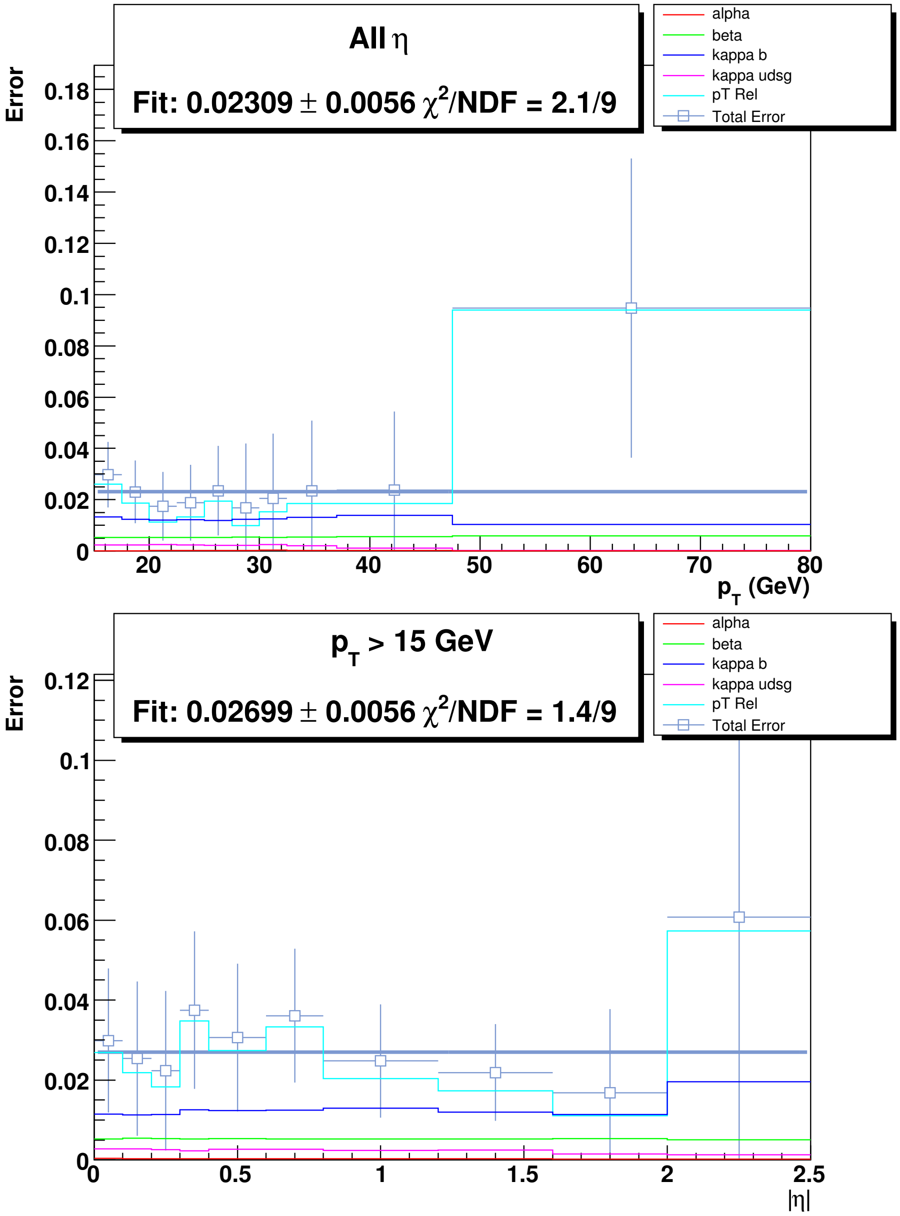

Figure 4.10: The total systematic error from the S8 correlation factors parametrised in p_T and η and the fits taken as the overall systematic errors. The constituent errors for each of the correlation factors are also shown.

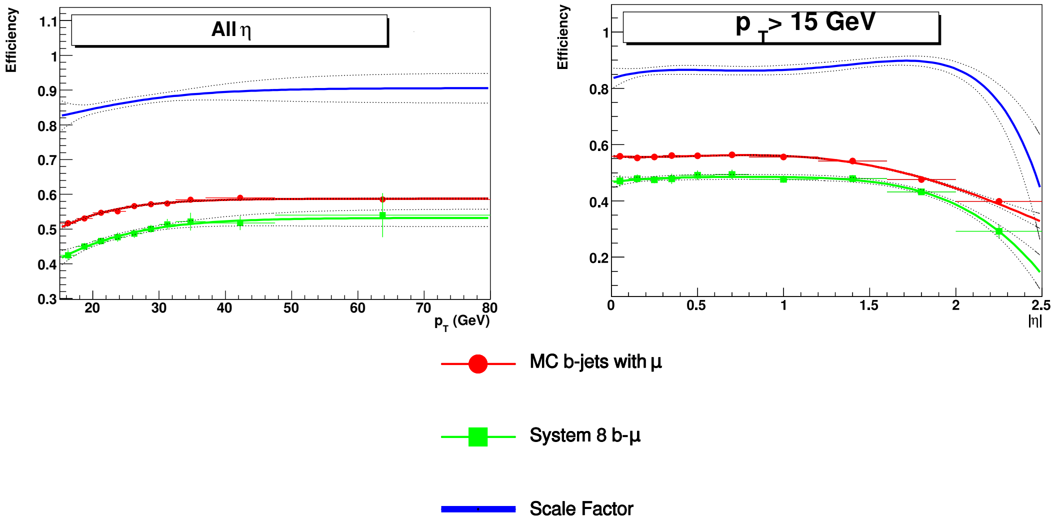

Figure 4.11: Tagging efficiency for muonic *b*-jets for data (green points) and MC (red points) with fits (green/red curves) and the scale factor (blue curve) derived from their ratio. The parametrisation in p_T is shown on the left and that in η on the right. The dotted black curves represent the error bars on the fits. The errors on the data efficiency derived from S8 include the errors due to the S8 coefficients. The scale factor errors are the sum in quadrature of the two efficiency errors and the systematic errors derived later in this Chapter.

4.3.3.2 Data/MC Scale Factor and Tag Rate Functions

The data/MC scale factor (SF) reflects differences in tracking between data and MC. The scale factor is the curve resulting from the division of the fit to the data muonic *b*-efficiency by the fit to the MC muonic *b*-efficiency, i.e. the ratio of the two TRFs. The *b*-efficiency derived from data along with the MC *b*-jet efficiency for muonic events and the SF for converting between the two are shown for the Tight NN operating point in Figure 4.11.

The SF is expected to be valid for all samples such that it can be used to correct the *b*-efficiency measured on any MC sample to the true data efficiency.

The SF is thus used to convert the inclusive *b*-jet MC efficiency to give the inclusive *b*-jet data efficiency such that we have a valid tag rate function for all data samples, not only muonic ones. The inclusive MC *b*-efficiency and the inclusive *b*-jet data TRF are shown in Figure 4.12 for the Tight NN operating point. Similarly the tag rate on an inclusive MC *c*-jet sample is combined with the SF to give a *c*-jet TRF for data. Figure 4.13 shows the inclusive *c*-jet data TRF and the inclusive MC *c*-jet efficiency.

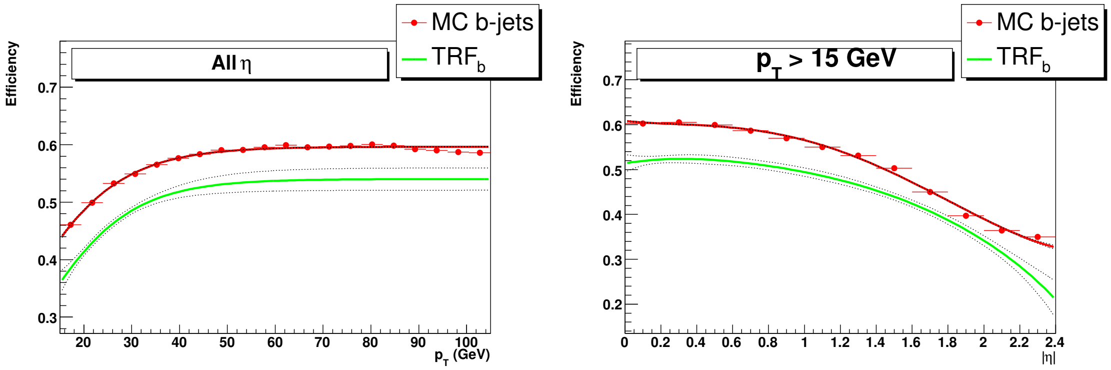

Figure 4.12: The inclusive MC *b*-efficiency measured (red points) and fit (red line) and the *b*-quark TRF (green curve) for the Tight operating point. The dotted lines give the error on the fit and the full errors on TRF_b.

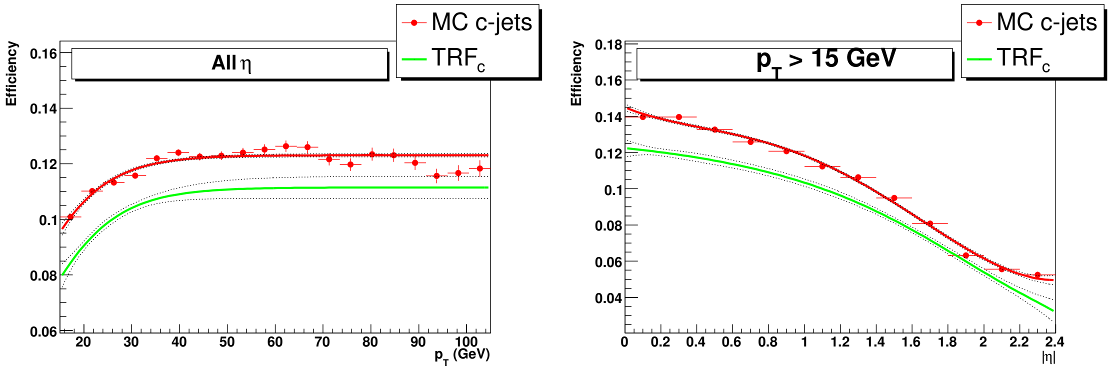

Figure 4.13: The inclusive MC *c*-efficiency measured (red points) and fit (red line) and the *c*-quark TRF (green curve) for the Tight operating point. The dotted lines give the error on the fit and the full errors on TRF_c.

4.3.4 Uncertainties

Two main sources of systematic errors need to be considered; the parametrisation error and the sample dependencies. The parametrisation of the efficiency into one-dimensional η and p_T functions inherently neglects correlations between p_T and η. In addition imperfections exist in the fits themselves, together these give the first source of systematics. Secondly it is acknowledged that different tagging efficiencies can be expected on different jet samples. Within the combined sets of b, $b \rightarrow \mu$ and c-jets the variation in efficiency between constituent samples is the second source of systematic error.

4.3.4.1 Closure Tests

The two sources of systematic error mentioned above can be considered together by performing closure tests on the constituent and combined MC samples. The number of tags predicted with the TRF derived on the combined MC samples is compared to the number of actual tags in an individual sample. More precisely the percentage difference is calculated, weighted by the fraction of events in the particular p_T-η bin. The root mean square (RMS) of the resulting distribution is the systematic error due to the individual sample.

Example closure test plots are shown in Figure 4.14 for the direct $b\bar{b}$ and $t\bar{t}$ samples compared to the inclusive b TRF. Closure tests are also performed on the combined samples used to derive the TRFs, giving a measure of the error due to the assumptions inherent in the parametrisation. Tables 4.9, 4.10 and 4.11 give the RMS errors for the closure tests on the inclusive b, muonic b and c-jet sub-samples.

The RMS error from the sub-sample with the largest deviation from closure was conservatively taken as the overall systematic error and carried forward to table 4.12. The closure tests gave a larger total uncertainty than the parametrisation and sample dependence errors when measured separately [72].

4.3.4.2 Systematic Uncertainties

Table 4.12 gives the final systematic error values for each jet type and the error assigned to the SF and TRFs. The systematic uncertainties are assigned as follows:

SF error - The RMS closure test error for the muonic b-jet samples (Table 4.10).

TRF$_b$ error - The systematic on the SF added in quadrature to the RMS error for the inclusive b-jet samples (Table 4.9).

TRF$_c$ error - The systematic on the SF added in quadrature to the RMS error for the c-jet samples (Table 4.11).

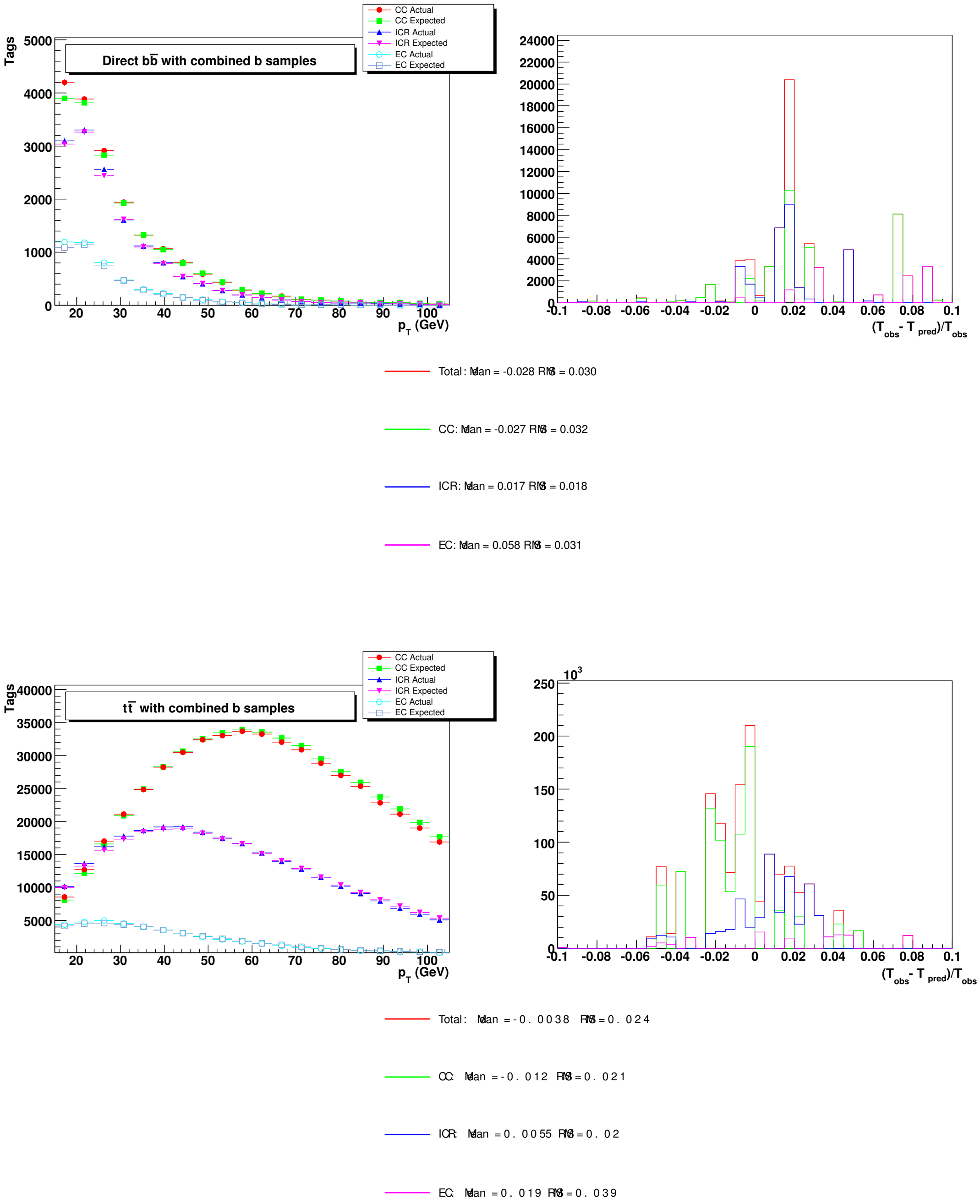

Figure 4.14: Closure tests on direct $b\bar{b}$ (top) and $t\bar{t}$ (bottom) samples for the three η regions. Left plots show the actual number of tags on the given MC sample and the expected tags predicted by the parametrised MC efficiency measurement. Right shows the normalised difference between observed (T_{obs}) and expected (T_{pred}) results. The RMS of this distribution gives the RMS closure test systematic error.

RMS	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
Combined b	1.14%	1.28%	1.43%	1.58%	1.79%	2.11%
QCD $b\bar{b}$	1.19%	1.22%	1.22%	1.20%	1.32%	1.68%
$t\bar{t}$	0.95%	1.03%	1.19%	1.33%	1.51%	1.79%
$\mathrm{Z}^0 \rightarrow b\bar{b}$	1.16%	1.19%	1.24%	1.33%	1.44%	1.54%
Max	1.19%	1.28%	1.43%	1.58%	1.79%	2.11%
RMS	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
Combined b	2.20%	2.54%	2.84%	3.17%	3.45%	3.45%
QCD $b\bar{b}$	1.63%	1.64%	1.76%	2.09%	2.36%	2.45%
$t\bar{t}$	1.84%	2.19%	2.43%	2.74%	2.98%	3.14%
$\mathrm{Z}^0 \rightarrow b\bar{b}$	1.59%	1.95%	2.01%	2.07%	2.4%	2.36%
Max	2.20%	2.54%	2.84%	3.17%	3.45%	3.45%

Table 4.9: The RMS errors from the b-quark closure tests, for the individual and combined samples. The largest is taken as the error on the TRF.

RMS	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
Combined $b \rightarrow \mu$	1.20%	1.44%	1.63%	1.82%	1.96%	2.33%
QCD $b\bar{b} \rightarrow \mu$	0.30%	0.22%	0.17%	0.32%	0.47%	0.84%
$t\bar{t} \rightarrow b \rightarrow \mu$	1.29%	1.45%	1.59%	1.70%	1.83%	2.13%
$\mathrm{Z}^0 \rightarrow b\bar{b} \rightarrow \mu$	1.44%	1.77%	1.98%	2.12%	2.39%	2.92%
Max	1.44%	1.77%	1.98%	2.12%	2.39%	2.92%
RMS	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
Combined $b \rightarrow \mu$	2.38%	2.64%	2.81%	2.86%	3.15%	3.35%
QCD $b\bar{b} \rightarrow \mu$	1.03%	1.50%	1.94%	2.39%	2.67%	2.55%
$t\bar{t} \rightarrow b \rightarrow \mu$	2.27%	2.35%	2.61%	2.66%	2.92%	3.05%
$\mathrm{Z}^0 \rightarrow b\bar{b} \rightarrow \mu$	3.08%	3.51%	3.47%	3.58%	3.36%	3.09%
Max	3.08%	3.51%	3.47%	3.58%	3.36%	3.35%

Table 4.10: The RMS errors from the muonic b-jet closure tests, for the individual and combined samples. The largest is taken as the error on the TRF.

RMS	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
Combined c	2.00%	2.10%	2.35%	2.56%	2.92%	3.25%
QCD $c\bar{c}$	2.39%	1.80%	1.95%	2.04%	2.17%	2.97%
$t\bar{t} \rightarrow c \rightarrow \mu$	2.01%	2.09%	2.36%	2.52%	2.71%	2.95%
$\mathrm{Z}^0 \rightarrow c\bar{c}$	1.99%	2.04%	2.27%	2.67%	2.86%	2.99%
Max	2.39%	2.10%	2.36%	2.67%	2.92%	3.25%

RMS	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
Combined c	3.48%	3.81%	3.78%	3.31%	3.56%	3.55%
QCD $c\bar{c}$	2.90%	3.16%	2.50%	3.73%	2.97%	3.96%
$t\bar{t} \rightarrow c \rightarrow \mu$	3.25%	3.27%	3.73%	4.17%	3.81%	3.80%
$\mathrm{Z}^0 \rightarrow c\bar{c}$	2.95%	3.18%	3.88%	3.88%	3.87%	4.03%
Max	3.48%	3.81%	3.88%	4.17%	3.87%	4.03%

Table 4.11: The RMS errors from the c-quark closure tests, for the individual and combined samples. The largest is taken as the error on the TRF.

Error	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
MC $b \rightarrow \mu$	1.44%	1.77%	1.98%	2.12%	2.39%	2.92%
MC b	1.19%	1.28%	1.43%	1.58%	1.79%	2.11%
MC c	2.39%	2.10%	2.36%	2.67%	2.92%	3.25%
SF	1.44%	1.77%	1.98%	2.12%	2.39%	2.92%
TRF_b	1.87%	2.18%	2.44%	2.64%	2.98%	3.60%
TRF_c	2.79%	2.74%	3.08%	3.41%	3.77%	4.37%

Error	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
MC $b \rightarrow \mu$	3.08%	3.51%	3.47%	3.58%	3.36%	3.35%
MC b	2.20%	2.54%	2.84%	3.17%	3.45%	3.45%
MC c	3.48%	3.81%	3.88%	4.17%	3.87%	4.03%
SF	3.08%	3.51%	3.47%	3.58%	3.36%	3.35%
TRF_b	3.79%	4.34%	4.48%	4.78%	4.81%	4.81%
TRF_c	4.65%	5.18%	5.20%	5.50%	5.13%	5.24%

Table 4.12: Total systematic errors on the SF and the TRFs from the MC sample parametrisations.

4.3.4.3 Statistical Error

The statistical error is given by:

$$\sigma_{stat} = \frac{f^{\pm\sigma}(p_T) \times f^{\pm\sigma}(\eta)}{f^{\pm\sigma}_{total}} - \frac{f(p_T) \times f(\eta)}{f_{total}} \tag{4.19}$$

Where f are the TRF or SF parametrisations and f_{total} is the overall mean required for normalisation. The TRF and SF parametrisations are varied within the fit errors, $\pm\sigma$, and the overall efficiency is recalculated. The statistical error, σ_{stat}, is taken as the maximum difference between this new efficiency (the first term in Eqn. 4.19) and the efficiency calculated with the central value fits (the second term).

4.3.4.4 Total Error

The systematic and statistical errors are combined in quadrature to give the total error for each p_T or η bin and each operating point. Figure 4.15 shows the total error for the SF, TRF$_b$ and TRF$_c$. At low p_T and high η statistical errors become dominant and the error curves depart from the constant total systematics.

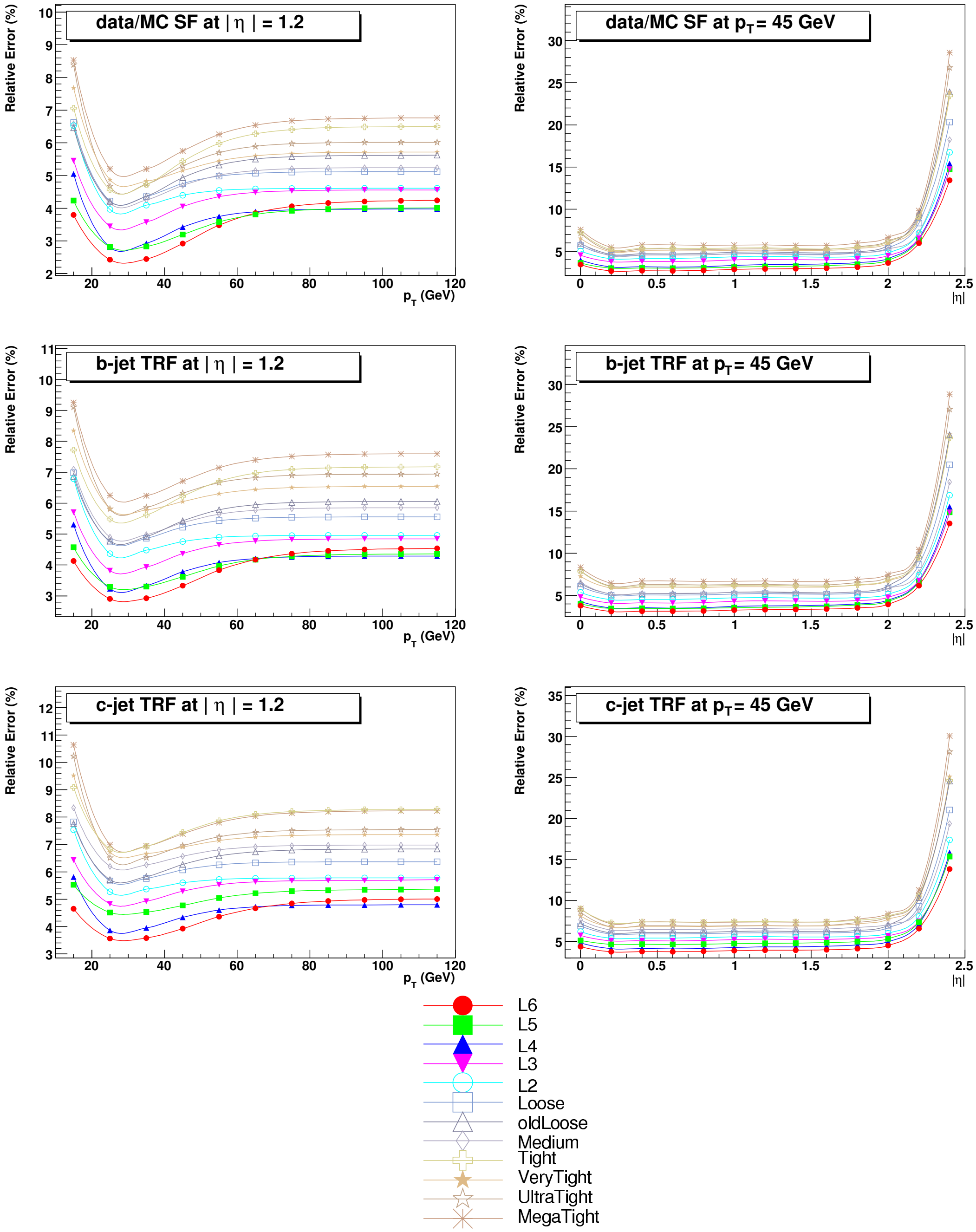

Figure 4.15: Total errors on the SF (top), TRF$_b$ (middle) and TRF$_c$ (bottom) for the twelve operating points parametrised in p_T (left) and η (right).

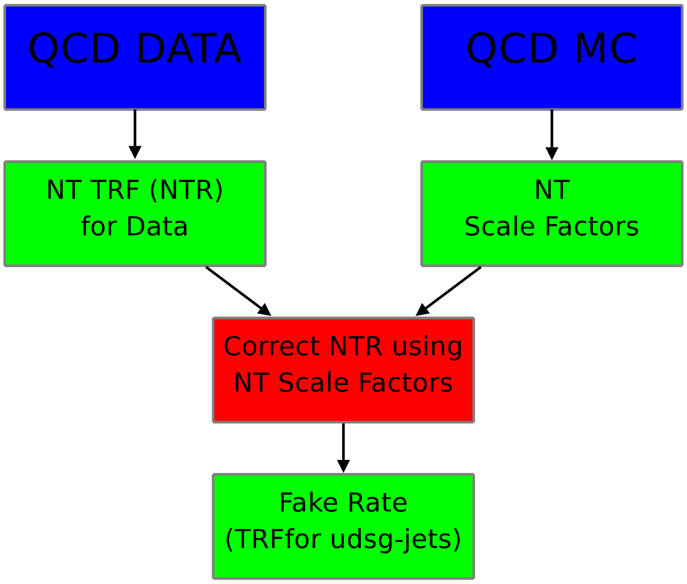

Figure 4.16: Schematic describing the process of calculating the fake tag rate.

4.4 Data Certification for Fake Rate

Figure 4.16 gives an overview of the procedure used to measure the fake rate from data using the negative tag rate and correction factors from MC.

4.4.1 Negative Tag Rate

The negative tag rate (NTR) was calculated using the EM and QCD skims combined (COMB skim), as a first estimate of the data fake rate[1]. For fake jets a different p_T spectrum was seen in the different calorimeter η regions and so the factorisation of the tag rate used for the b-efficiency was not possible. The sample of jets was thus divided into the three η regions described in Section 4.1.2.4. Given the more limited statistics it was only possible to parametrise with respect to p_T within each η region.

The negative tag rates for the COMB skim are shown integrated over p_T for all operating points for the three η regions in Table 4.13. The negative tag rate measured in p_T bins in the three η regions is shown in Figure 4.17 for the COMB skim as well as for the constituent EM and QCD skims. The negative tag rate functions are second order polynomials fit to these distributions.

---

[1] Jets close to EM clusters were removed - see Section 4.4.3.

CAL Region	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
CC	12.40%	9.30%	7.02%	5.49%	3.97%	2.51%
ICR	11.20%	8.30%	6.16%	4.74%	3.35%	2.05%
EC	10.40%	7.60%	5.44%	4.03%	2.73%	1.55%
CAL Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	2.11%	1.23%	0.72%	0.47%	0.32%	0.25%
ICR	1.70%	0.96%	0.53%	0.34%	0.23%	0.17%
EC	1.25%	0.63%	0.30%	0.17%	0.10%	0.07%

Table 4.13: Negative tag rates for the COMB skim.

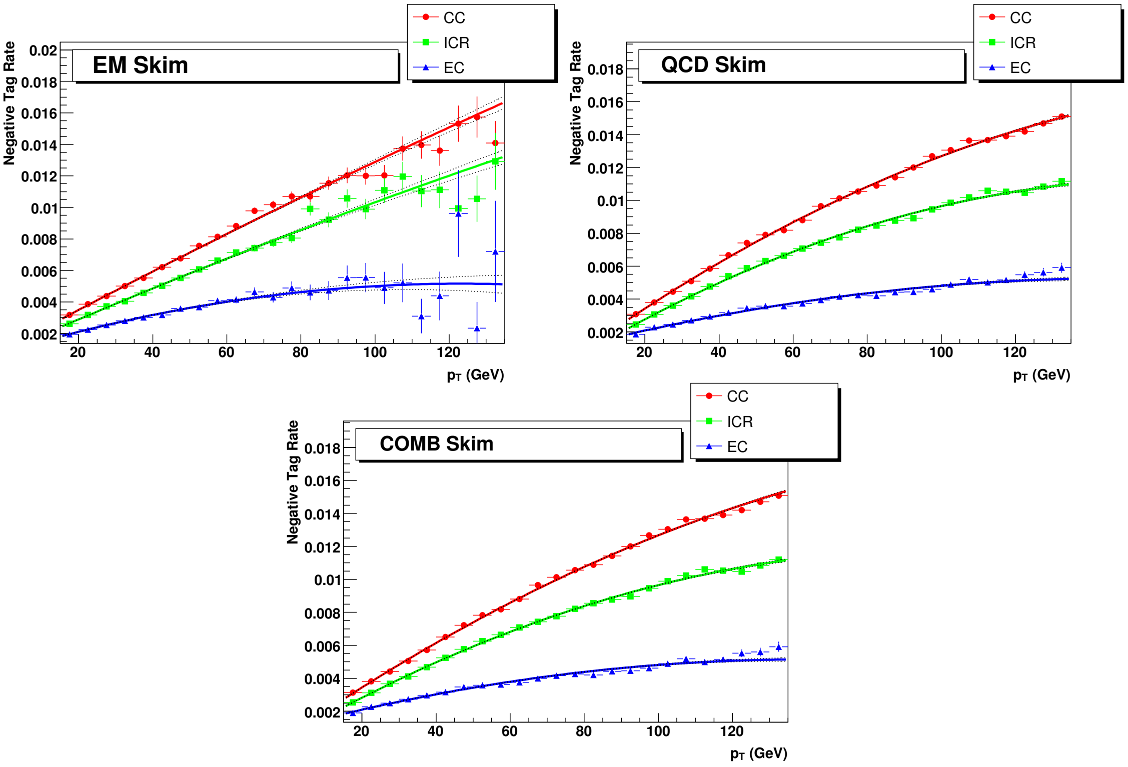

Figure 4.17: The negative tag rate measured on the EM skim (top left), QCD skim (top right) and the COMB skim (bottom) shown against p_T for the three η regions. The negative tag rate functions are the second order polynomial fits and are also shown.

Region	L6	L5	L4	L3	L2	Loose
CC	11.10%	8.16%	6.06%	4.66%	3.28%	2.02%
ICR	10.80%	7.93%	5.86%	4.48%	3.12%	1.84%
EC	10.90%	8.08%	5.92%	4.35%	2.98%	1.69%
Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	1.68%	0.96%	0.55%	0.34%	0.23%	0.17%
ICR	1.50%	0.80%	0.41%	0.24%	0.17%	0.13%
EC	1.35%	0.66%	0.30%	0.17%	0.09%	0.05%

Table 4.14: The FTR calculated from the NTR on the COMB data skim corrected for heavy flavour and the positive-negative tag asymmetry.

4.4.2 Converting Negative Tag to Fake Rate

Resolution is not the only effect which contributes to the fake rate, in addition the negative tag rate is not quite as immune to signal contamination as has been suggested so far. To correct the negative tag rate to find the actual data fake rate we need to account for these factors.

Since the b- and c-quark mean lifetimes are of a similar order of magnitude to the resolution width it is possible that some heavy flavour quark decays may produce tracks which appear in the negative tags. This will manifest itself as an increased fake tag rate which we remove using the heavy flavour scale factor, SF_{hf}. Some light flavour composite states such as K_S^0, K_L^0, Λ and γ conversions are also long lived and will increase the positive tag rate (PTR) over the negative tag rate. This boost to the PTR is accounted for by SF_{lf} the light jet PTR/NTR ratio. In all the fake rate is calculated using:

$$FTR = SF_{hf} \times SF_{lf} \times NTR \tag{4.20}$$

A QCD MC sample with a wide p_T range was used to estimate the two scale factors as shown in Figure 4.18. The product of the two factors as a function of p_T was calculated and fit with a constant. The fit errors were scaled up by the $\sqrt{\chi^2/NDF}$ to account for the uncertainty in the parametrisations. Figure 4.18 also shows the total correction factor fit for MC.

The fake rate on the COMB skim for the Tight NN operating point can be seen in Figure 4.19. The p_T integrated rate for all operating points for the three η regions is given in Table 4.14.

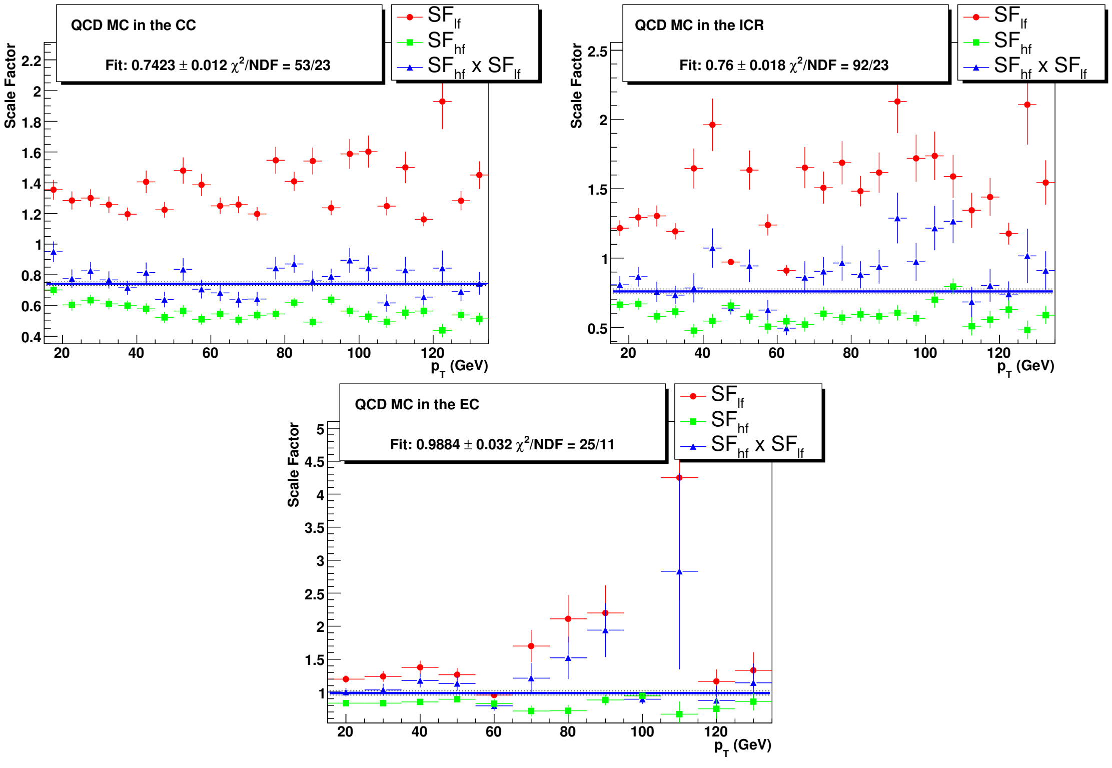

Figure 4.18: The positive-negative tag asymmetry correction (SF_{lf}), heavy flavour correction (SF_{hf}) and the fit to the overall correction factor (the product of the two).

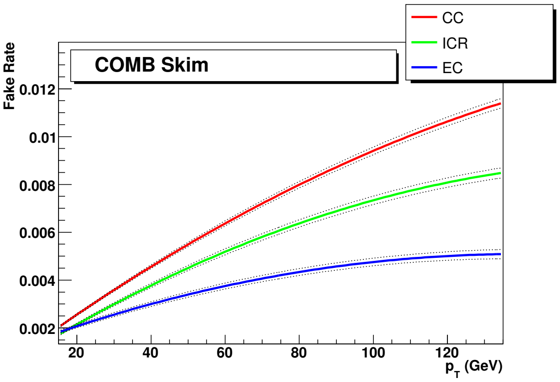

Figure 4.19: The FTR parametrised with respect to p_T for the three η regions. The dotted lines give the total errors on the FTR (Section 4.4.3).

Region	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
CC	0.13%	0.15%	0.17%	0.19%	0.24%	0.30%
ICR	0.04%	0.04%	0.05%	0.06%	0.07%	0.09%
EC	0.05%	0.06%	0.08%	0.11%	0.15%	0.21%
Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	0.32%	0.38%	0.47%	0.52%	0.57%	0.60%
ICR	0.09%	0.12%	0.17%	0.21%	0.25%	0.29%
EC	0.23%	0.37%	0.51%	0.54%	0.60%	0.72%

Table 4.15: Error assigned due to the EM and QCD skim NTR difference.

4.4.3 Systematic Uncertainties

4.4.3.1 Negative Tag Sample Dependence Systematic

Different negative tag rates are found when considering the EM and QCD samples separately. The EM sample has an enhanced fake rate from electrons and photons. If jets close to EM clusters are removed the two samples become similar enough that they can be combined to form the COMB skim and a NTR consistent with both subsamples can be calculated. The remaining difference between the NTRs on the QCD and EM samples is used to give a systematic error due to the sample composition. The EM/QCD NTR ratio is fit with a constant and half the difference from unity is used as the systematic. For some operating points the fit is consistent with unity within its error, in this case the fit error scaled by $\sqrt{\chi^2/NDF}$ is used as the systematic. The errors for the various operating points are summarised in Table 4.15.

Some analysis groups were uncomfortable with the removal of jets near EM clusters since this was expected to impact their signal efficiency. Assuming the analysis triggers do not select EM objects these jets will be rare and the effect of allowing jets near EM clusters will be small. The ratio of the negative tag rates measured on the QCD skim with and without jets near EM clusters was constructed and systematic errors calculated as with the EM-QCD difference above. The ratios in the three η regions and resulting errors can be seen in Figure 4.20 and Table 4.16. The effect is small and p_T independent, ranging from 0.2% to 1%. The inclusion of this error allows analysers the choice of using the fake rate either in a low EM sample without EM removal or in any sample with EM removal. The effect of applying this EM cluster jet removal to the BID skim used for the b-efficiency calculation was also examined and found to be negligible.

Region	L6	L5	L4	L3	L2	Loose
CC	0.19%	0.24%	0.28%	0.33%	0.37%	0.46%
ICR	0.17%	0.19%	0.25%	0.28%	0.29%	0.39%
EC	0.11%	0.16%	0.21%	0.26%	0.30%	0.39%
Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	0.49%	0.62%	0.72%	0.78%	0.87%	0.95%
ICR	0.45%	0.55%	0.67%	0.76%	0.82%	0.86%
EC	0.39%	0.49%	0.63%	0.58%	0.67%	0.74%

Table 4.16: Error arising from the difference between NTRs derived on the QCD skim with and without the removal of jets close to EM clusters, measured at the Tight NN operating point.

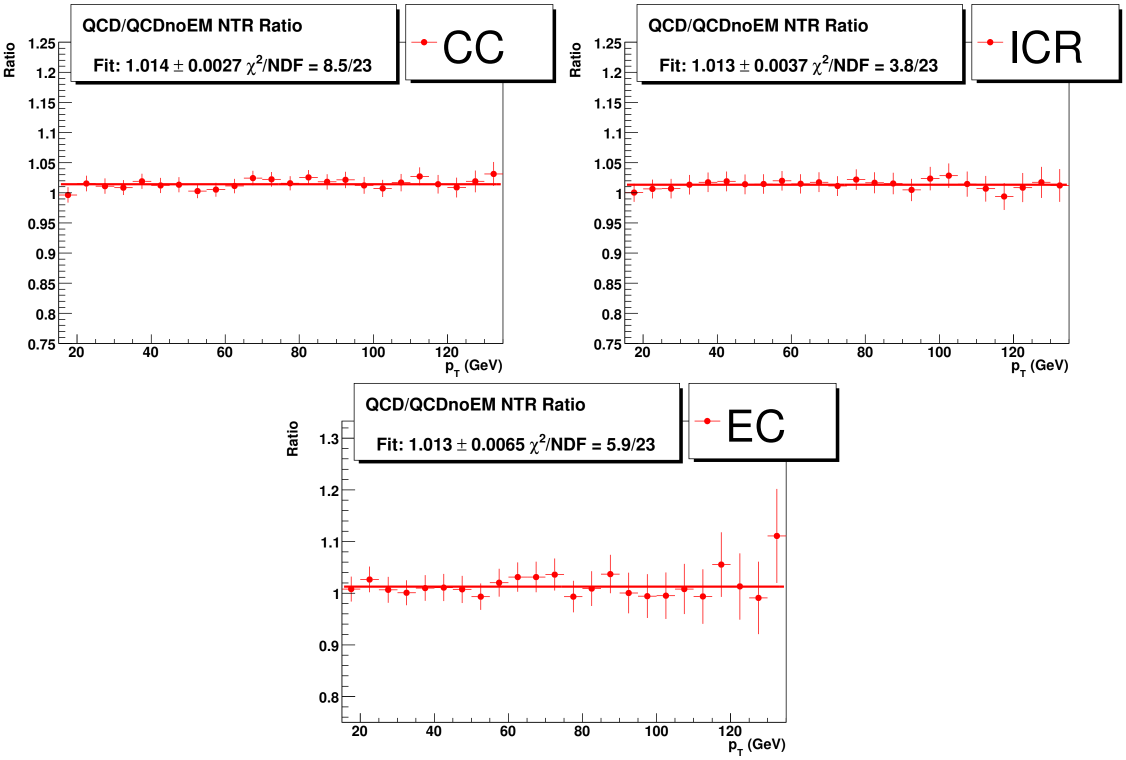

Figure 4.20: The fractional difference between NTRs derived on the QCD skim with and without the removal of jets close to EM clusters, measured at the Tight NN operating point.

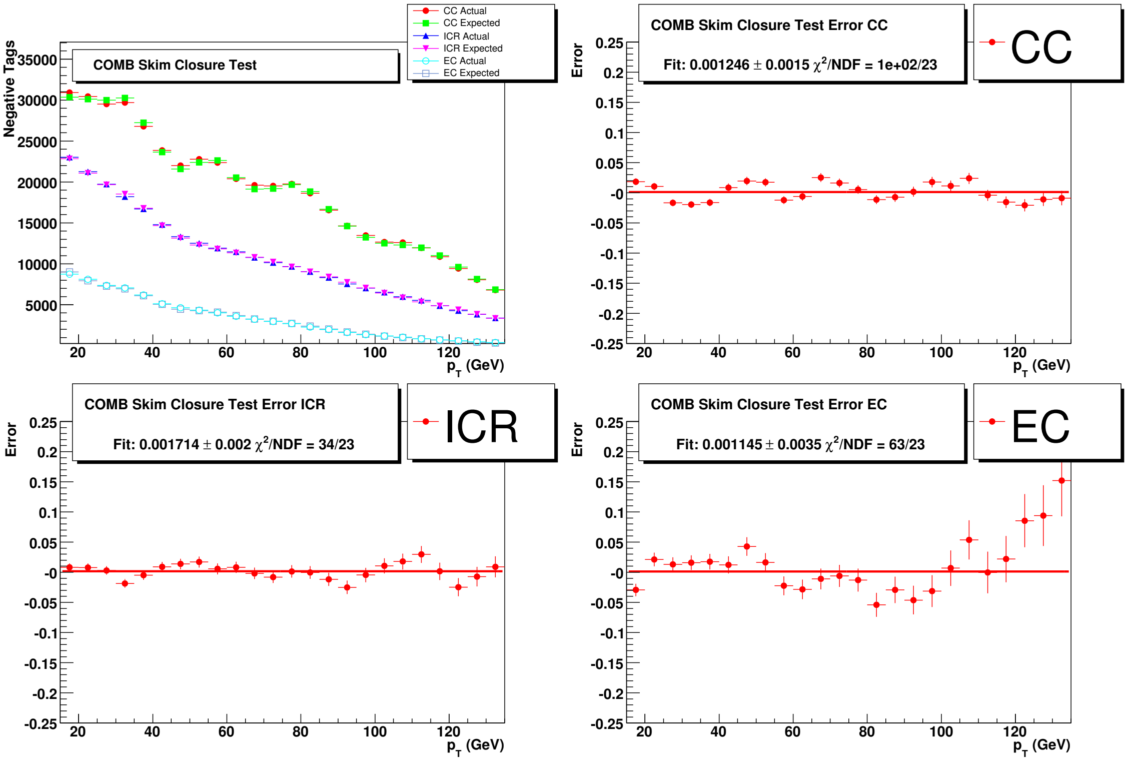

Figure 4.21: The closure test on the COMB skim for the Tight NN operating point in the three η regions, top left. The fractional difference between predicted and actual tags is also shown for each of the three regions.

4.4.3.2 Parametrisation Error

Closure tests like those for the b-efficiency were also performed. The NTR parametrisation was used to predict the number of negative tags in each η region as a function of p_T and this was compared to the actual number of negative tags. Figure 4.21 shows the actual and predicted tags for the COMB skim for each of the η regions and the ratios of actual to predicted tags by region. The systematic error due to the parametrisation of the NTR as a polynomial was taken to be the deviation from unity of a constant fit to these ratios. Where the fit was consistent with unity within the fit error scaled by $\sqrt{\chi^2/NDF}$, the scaled error was used instead. The resulting errors for each operating point are given in Table 4.17.

4.4.3.3 Correction Factor Error

The heavy flavour scale factor depends on the b- and c-quark content of the sample. Since the heavy flavour content in the EM and QCD samples is not well known it is unclear what level to use in the MC sample used to simulate SF_{hf}. The number of b- and c-quarks present was varied by ±20% and the fake rate was recalculated. The

Region	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
CC	0.06%	0.06%	0.07%	0.07%	0.08%	0.09%
ICR	0.07%	0.08%	0.09%	0.09%	0.11%	0.13%
EC	0.06%	0.07%	0.09%	0.10%	0.12%	0.16%

Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	0.10%	0.11%	0.15%	0.19%	0.22%	0.25%
ICR	0.13%	0.15%	0.20%	0.25%	0.30%	0.35%
EC	0.17%	0.25%	0.35%	0.47%	0.61%	0.72%

Table 4.17: The systematic uncertainty arising from the parametrisation of the NTR on the COMB skim.

largest change in fake rate was taken as the systematic error due to the uncertain *b*- or *c*-quark content. The errors due to *b*-quark content are given in Table 4.18 and for *c*-quarks in 4.19.

Region	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
CC	1.96%	2.24%	2.57%	2.88%	3.25%	3.84%
ICR	1.56%	1.78%	2.14%	2.45%	2.80%	3.22%
EC	0.68%	0.76%	0.86%	0.93%	1.08%	1.34%

Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	4.05%	4.94%	5.64%	6.08%	6.45%	6.67%
ICR	3.38%	4.10%	5.09%	5.65%	6.73%	7.07%
EC	1.38%	1.43%	1.55%	2.60%	3.68%	3.41%

Table 4.18: The systematic due to uncertainty in the required *b*-quark content of the MC.

4.4.3.4 Total Systematic Error

The systematic errors from the *b*- and *c*-content are likely to be correlated and are therefore added directly. The remaining errors, from the parametrisation and sample content, are added in quadrature. Table 4.20 gives the combined systematic uncertainties. These values include both the EM/QCD comparison error and the error from considering the QCD sample with and without jets near EM clusters.

4.4.3.5 Combined Systematic and Statistical Error

The total error was calculated by adding in quadrature the total systematic errors with the statistical errors, where the statistical errors were given by $\pm 1\sigma$ curves on

Region	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
CC	1.74%	1.89%	2.06%	2.22%	2.35%	2.48%
ICR	1.46%	1.60%	1.77%	1.88%	2.01%	2.15%
EC	0.87%	0.93%	1.04%	1.02%	1.03%	1.07%
Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	2.48%	2.63%	2.66%	2.56%	2.48%	2.39%
ICR	2.22%	2.46%	2.40%	2.44%	2.28%	2.40%
EC	1.04%	1.14%	0.87%	1.37%	0.96%	0.94%

Table 4.19: The systematic due to uncertainty in the required c-quark content of the MC.

Region	**L6**	**L5**	**L4**	**L3**	**L2**	**Loose**
CC	3.71%	4.15%	4.65%	5.13%	5.64%	6.36%
ICR	3.04%	3.42%	3.98%	4.40%	4.85%	5.44%
EC	1.99%	2.17%	2.44%	2.50%	2.72%	3.01%
Region	**oldLoose**	**Medium**	**Tight**	**VeryTight**	**UltraTight**	**MegaTight**
CC	6.59%	7.68%	8.41%	8.73%	9.05%	9.23%
ICR	5.66%	6.60%	7.54%	8.14%	9.07%	9.53%
EC	3.07%	3.39%	3.10%	4.76%	5.71%	5.60%

Table 4.20: Total FTR systematic errors. These include the QCD/EM sample uncertainty and the QCD skim EM cluster removal uncertainty.

the FTR parametrisation fit. The systematic errors are generally dominant with statistical errors having an increasing contribution for the tighter operating points. In Figure 4.19, which shows the parametrised fake tag rates, the total errors are given as dotted lines.

4.5 NN Tagger Performance

The performance of the NN tagger measured on data with full systematic and statistical errors is shown in Figure 4.22. The performance of the individual JLIP tagger, which was previously the most widely used tagger, is shown for comparison. The NN consistently shows a greater efficiency for a given fake rate across the full range of operating points. The performance on MC is also shown for Z $\rightarrow b\bar{b}/q\bar{q}$ and for QCD b/uds jets in Figure 4.23. This is the uncorrected MC performance, calculated by tagging MC directly. It is clear that for JLIP the performance in MC and data are quite different, this is the difference that makes the complex certification procedure and the calculation of data/MC SFs necessary. For the NN this difference is smaller as is expected since the NN SF is a largely flat function and close to unity. It can be concluded that the NN is less sensitive to data/MC differences.

4.6 Conclusions

The latest version of the NN *b*-tagger has been certified for use at DØ. This work of training, testing and parametrising the tool has been reviewed and approved by the Collaboration. For the first time the NN tagger has become the default tagger for all *b*-tagging at the experiment and so is very widely used. The number of NN operating points made available to analysers was increased from six to twelve. The measurements and parametrisations described above for the NN Tight operating point were repeated for each operating point. The S8 correlation coefficient α and error were measured on MC for the first time. Previously a constant was assigned to α and a large error was required, measuring α allowed a significant reduction in the overall *b*-efficiency error. The NN certification code was also modified to produce a more generic certification tool, which was officially made available to the Collaboration. This code was subsequently used to produce the Run IIb certification. Documentation for the certification tools was also developed.

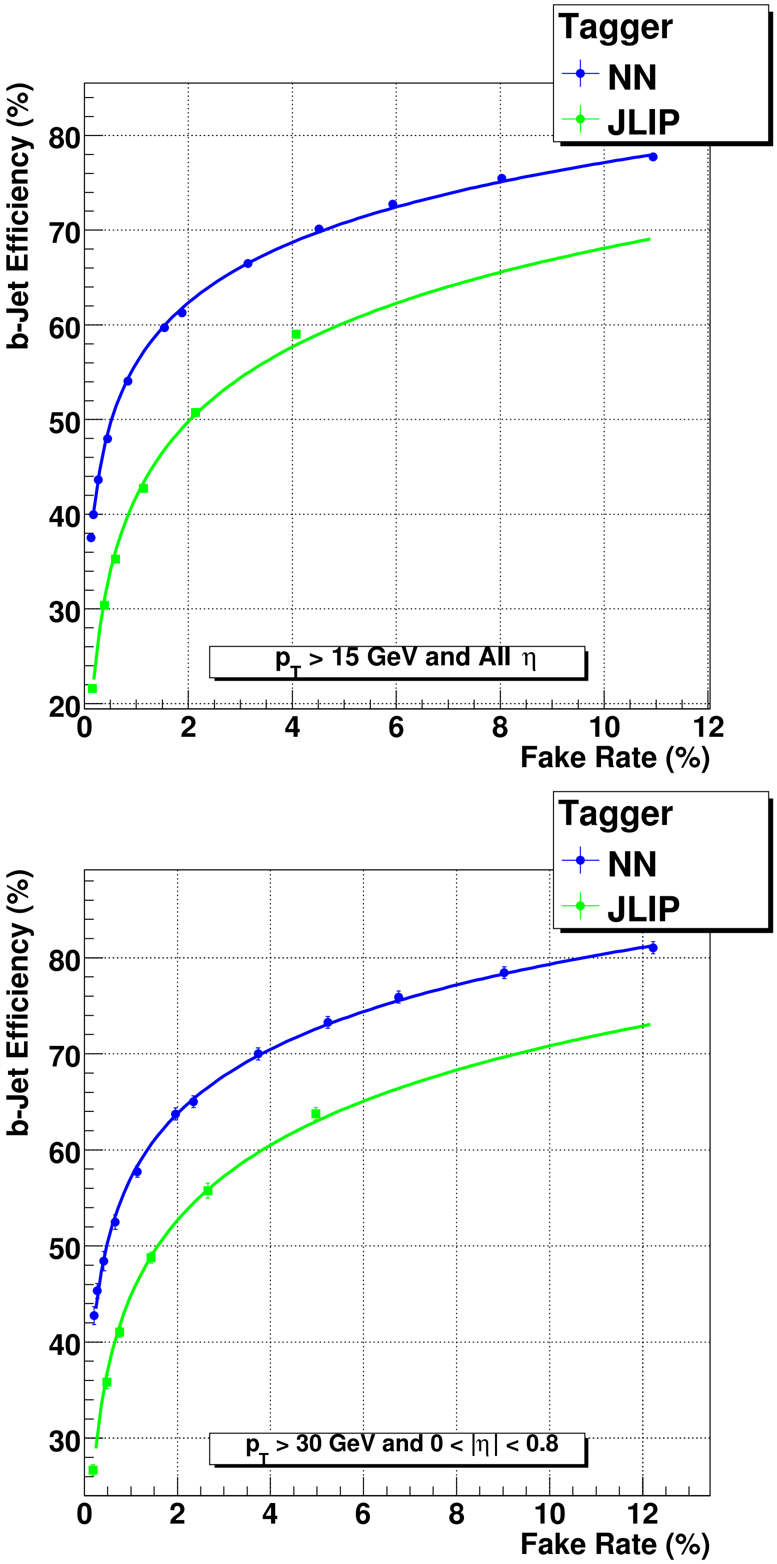

Figure 4.22: B-tagging performance for the NN- and JLIP-taggers on data. Upper plot for all taggable jets. Lower plot for higher p_T central jets ($p_T > 30$ GeV and $|\eta| < 0.8$).

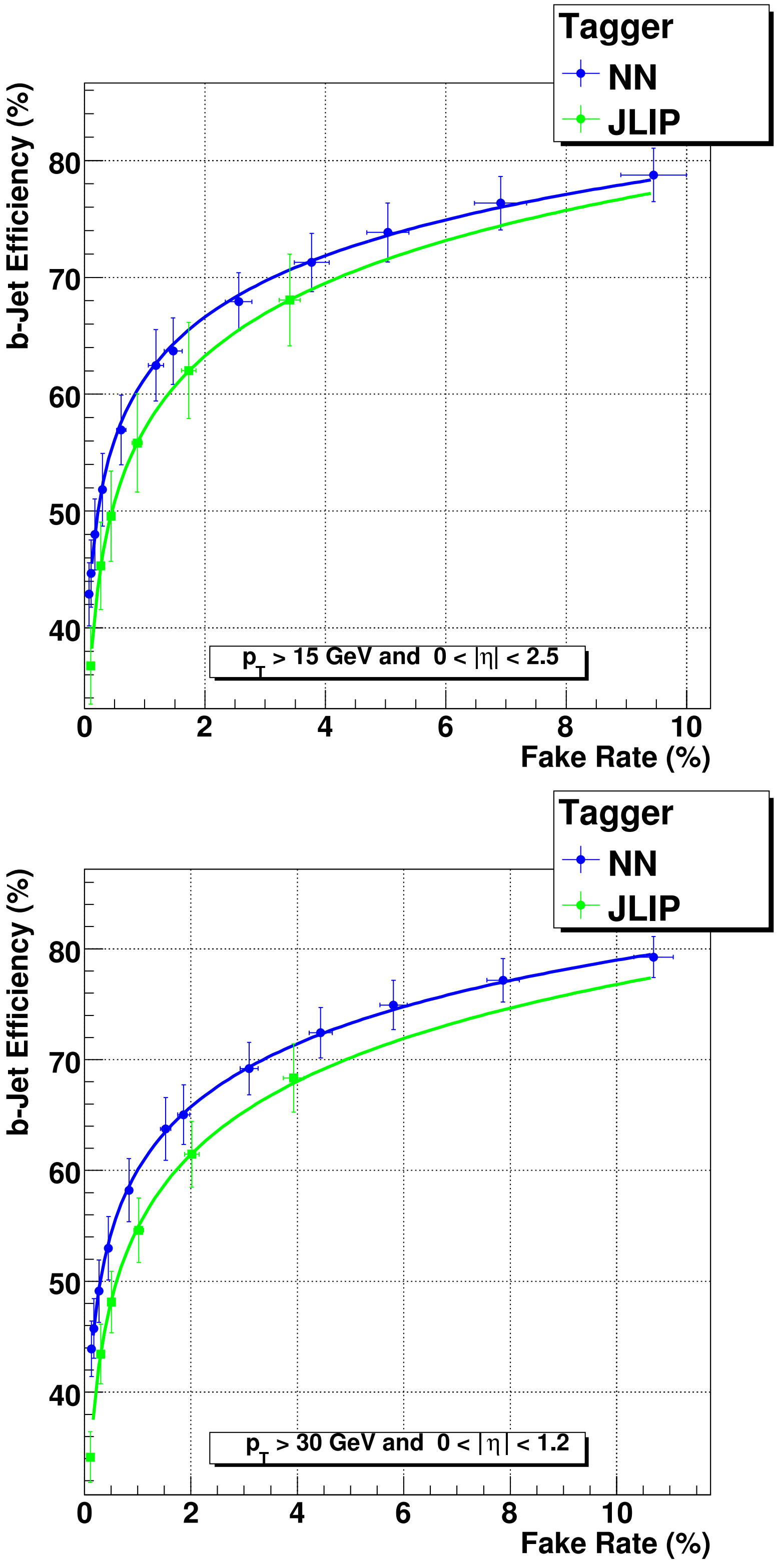

Figure 4.23: B-tagging performance for the NN- and JLIP-taggers on MC using the data/MC corrections. Upper plot shows the performance for $t\bar{t}$ and $Z \to q\bar{q}$. Lower plot shows direct $b\bar{b}$ with QCD light jets.

Chapter 5

The SUSY Higgs analysis

5.1 Introduction

The SUSY Higgs analyses are important search channels for the Tevatron experiments. These analyses offer the possibility of answering two of the most pressing questions in particle physics, namely to discover the Higgs boson and to find indications of new physics beyond the SM. One of the major channels is the $bh \rightarrow bbb$ topology as discussed in Section 1.3. The following Chapter discusses the preliminary result shown at ICHEP 2006 using $\sim 1fb^{-1}$ of data and, in greater detail, the improved version submitted for publication. This analysis relies very heavily on b-tagging, both online and offline, and thus makes extensive use of the work discussed in Chapters 3 and 4. It was the first analysis to use NN b-tagging at a hadron collider. As well as the generic b-tagging developed, necessary for this analysis, the author produced the taggability parametrisations and optimised the b-tagger working point for both analysis iterations. The submitted analysis uses a likelihood discriminant to improve the sensitivity. As an independent cross-check the author developed a neural network to be used as an alternative discriminant. Full descriptions of the preliminary and submitted results can be found in [24] and [27, 22] respectively. Section 5.2 gives an outline of the two analysis versions. The rest of this Chapter focusses on the submitted result. Sections 5.3 and 5.4 describe the data and MC samples used. Sections 5.5 and 5.6 discuss the background estimation and analysis optimisation. The development of the neural network discriminant is described in Section 5.7. Sections 5.8 and 5.9 cover the systematic errors and the limits set in the submitted result.

5.2 Analysis Overview

The basic event selection for the analysis requires three b-tagged jets with relatively high p_T, typically the jets from the Higgs candidate should have $p_T > 30$ GeV. The dominant backgrounds before and after b-tagging are due to multi-jet QCD events. Before b-tagging QCD processes involving three light-jets dominate. After b-tagging the background is dominated by the processes with two or three real b-jets. The cross-sections for these QCD processes are poorly known, thus a major difficulty for the analysis is to determine the level and composition of the background. The signal would be an excess of events over the background expectation peaked in the di-jet invariant mass distribution around the Higgs mass, m_h, in the sample with three b-tags. The more powerful online and offline b-tagging discussed in Chapters 3 and 4 has had a significant impact on the analysis, changing the background composition considerably since the previous published result [19]. The key difference between the two analysis iterations described here is the background estimation.

5.2.1 ICHEP 2006 Preliminary Result

The estimation of the background followed the procedure used in the previous published result [19]. The sample with three b-tags is known as the '3-tag sample' and that with two b-tags as the '2-tag sample'. The invariant mass of the two highest p_T jets is referred to as m_{01}. For the preliminary result the 3-tag background invariant mass distribution is produced from the 2-tag data distribution by weighting each 2-tag event with a $2 \rightarrow 3$-tag probability. This $2 \rightarrow 3$-tag probability is derived from data and is parametrised in terms of the p_T of the third jet to be b-tagged. To avoid any possible signal contamination it is derived outside the signal mass window and re-derived for each tested signal mass. This 3-tag background distribution, derived exclusively from data, is used as the background hypothesis in the limit setting. A MC background distribution is also produced to be used as a cross-check. The relative cross-sections of the MC background components are taken from the LO ALPGEN simulation. The total 2- and 3-tag MC distributions are then normalised to the data, again outside the signal region. Although the background components have similar m_{01} distributions the fact that the normalisation factors needed in the 2- or 3-tag MC samples are similar, though the sample composition changes, suggests that the relative levels of the most significant background components are reasonable. Figure 5.1 shows the agreement between the data and background-only hypothesis, along with the signal distribution and signal-plus-background hypothesis for a Higgs signal with $m_A = 170$ at the observed 95% confidence limit ($\tan\beta = 121$).

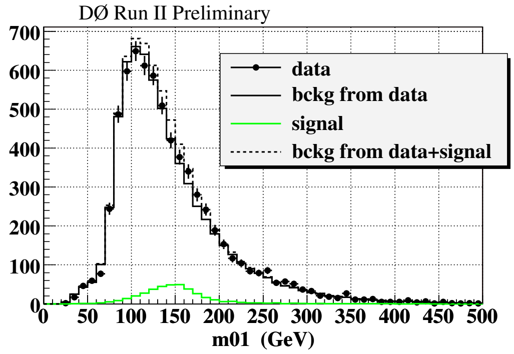

Figure 5.1: The 3-tag m_{01} distribution in the preliminary result [24] for data (circles), the background hypothesis derived from 2-tag data (solid black line), the signal MC (green line) for a Higgs with $m_A = 170$ and $\tan\beta = 121$ and the signal-plus-background hypothesis (dashed black line).

5.2.2 Final Publication

The NN b-tagger has considerably better performance than the SVT b-tagging used in the previous published result, thus after requiring three b-tags the background composition is very different in the two versions. As a result the three real b-jet component, which was previously small, is expected to make up a significant portion of the 3-tag background. This background component cannot be reduced relative to the signal by tighter b-tagging. With the background reduced relative to the signal due to the improved light-jet rejection the tails of the signal peak now become non-negligible across most of the mass spectrum. A signal-free or low-signal region with sufficient statistics to set the background level cannot be found. Thus an improved, more robust, method was developed. Instead the background composition is set using multiple correlated fits to the data H_T distribution with various numbers of b-tags, as described in Section 5.5. H_T is the scalar sum of jet p_Ts in the event. The expected background shape is produced from the 2-tag data which is background dominated and a 2 → 3-tag probability derived from a fit to the ratio of invariant mass distributions in 3-tag and 2-tag inclusive events in background-only MC. The overall normalisation of this predicted background shape is not determined and is allowed to float in the modified limit setting procedure. The result of not knowing the absolute background level is a weakening of the expected limits from the analysis. In order to recover sensitivity the jet p_T cuts are re-optimised and a multivariate discriminant is used. Since in data we do not know which jets originate from the Higgs the acceptance of the analysis can be maximised by considering multiple candidate di-jet pairs for each event.

5.2.3 Basic Event Selection

The basic event selection for the two versions of the analysis is the same; three high p_T b-jets with up to two additional jets from initial (ISR) and final state radiation (FSR). The following sections give a simple outline of the cuts used in the two versions.

5.2.3.1 Event Selection for the ICHEP Preliminary Result

To select signal events the following set of cuts was applied:

- The event must fire the $h^0 b\bar{b}$ trigger.
- Between 3 and 5 jets passing the 'good jet' criteria: $p_T > 15$ GeV, $|\eta| < 2.5$, taggable and passing the jet-ID selection.
- At least one jet above 45 GeV and one other jet above 25 GeV.
- At least 2 and 3 b-tagged jets in the 2-tag and 3-tag samples.
- The Higgs candidate is formed from the leading pair of jets.

5.2.3.2 Event Selection for the Publication

The improved event selection for the final publication requires:

- The event must fire the $h^0 b\bar{b}$ trigger.
- Between 3 and 5 jets passing the 'good jet' criteria: $p_T > 15$ GeV, $|\eta| < 2.5$, taggable and passing the jet-ID selection.
- All jets must have $p_T > 20$ GeV and at least two jets with $p_T > 25$ GeV.
- At least 2 and 3 b-tagged jets in the 2-tag and 3-tag samples.
- Multiple candidate di-jet pairs considered per event.
- The 3-, 4- and 5-jet events are considered as separate channels and optimised individually.
- A final selection using a multivariate discriminant.

5.3 Data

5.3.1 Triggers

Due to the limited bandwidth available for writing events to tape events are collected using a set of trigger requirements. For the search for Higgs bosons produced in association with b-quarks the $h^0b\bar{b}$ trigger has been developed. The data used in this analysis were collected with trigger versions v8 to v14, for these trigger versions the requirements of the $h^0b\bar{b}$ trigger are given in Table 5.1.

Over all its iterations the trigger's core requirement is for three jets with significant p_T. Since v11 the PV has been required to be within the silicon tracker barrel, increasing the likely taggability of jets. An IP-tagger term was introduced in v13 and requires that the probability of the event having no b-jets be less than 0.05. The majority of events ($\sim$ 70% of the integrated luminosity) were recorded using v13 and v14 which include this IP-tagger term; the IP-tagger is discussed in Chapter 3.

5.3.2 Skimming Cuts

The subset of data events to be considered by the analysis is called the skim. To be included in the skim events are required to pass the $h^0b\bar{b}$ trigger and have at least one jet with $p_T > 20$ GeV[1] and two others with $p_T > 15$ GeV. The jets must be within $|\eta| \leq 2.5$.

5.3.3 Data Quality

Blocks of luminosity marked as 'bad' by the DØ data quality group and runs with problems in the muon, calorimeter, SMT or CFT sub-systems were excluded from the data sample, as were events recorded during periods, within runs, with known calorimeter issues. After applying these requirements $1.0 fb^{-1}$ of data remains.

5.4 MC Samples

The MC samples were produced using Pythia [58] for the signal samples, $t\bar{t}$ and $Zb\bar{b}$. ALPGEN [63] was used for the multi-jet backgrounds containing heavy-flavour jets. Pythia is then used to model the showering of the ALPGEN partons into jets. The parton distribution functions (PDFs) used were those from CTEQ6L1 [73].

The events were then passed through the DØ detector simulation chain and overlaid with background elastic scattering events recorded with the 'zero bias' trigger,

---

[1] As measured before the jet energy scale (JES) correction is applied.

	v8	**v9**	**v10**
L1	CJT(3,7)	CJT(4,5,$\|\eta\| < 3.2$)	CJT(3,5,$\|\eta\| < 3.2$)
L2	-	JT(3,8,$\|\eta\| < 3$)HT(50,5)	same as v9
L3	JT(3,15,$\|\eta\| < 3$)	same as v8	same as v9
Name	3JT15	same as v8	same as v9

(a) v8, v9 and v10

	v11	**v12**
L1	CJT(3,5,$\|\eta\| < 3.2$)	same as v11
L2	JT(3,8,$\|\eta\| < 3$)HT(50,5)	same as v11
L3	JT(3,15,$\|\eta\| < 3$) $\times$ JT(2,25,$\|\eta\| < 3$) $\times$ $\|z_{\mathrm{PV}}\| < 35$ cm	JT(3,15,$\|\eta\| < 3$) $\times$ JT(2,25,$\|\eta\| < 3$) $\times$ $\|z_{\mathrm{PV}}\| < 35$ cm
Name	3JT15_PV	2J25_3J15_PVZ

(b) v11 and v12

	v13	**v14.0-7**	**v14.8-**
L1	same than v11 et v12	CJT(3,4,$\|\eta\| < 2.6$) $\times$ CJT(3,5,$\|\eta\| < 3.2$)	v14.0-7 $\times$ CJT(1,7,$\|\eta\| < 1.8$)
L2	JT(3,6,$\|\eta\| < 3$)HT(70,8)	same as v13	same as v13
L3	v12 $*$ $\mathrm{Prob_b}(0.05)$	same as v13	same as v13
Name	JT2_3JT15_IP_VX	same as v13	same as v13

(c) v13 and v14

Table 5.1: The triggering conditions for each version of the $h^0b\bar{b}$ trigger. The CJT(x,y,$|\eta| < z$) term corresponds to x calorimeter trigger towers above y GeV and within $|\eta| < z$. The JT(x,y,$|\eta| < z$) term corresponds to x jets reconstructed at Level-2 or Level-3 with $p_T > y$ GeV and $|\eta| < z$. The HT(x,y) term is used only at Level-2 and requires that the sum of the transverse momenta of L2 jets with $p_T > y$ GeV is above x GeV. The $\mathrm{Prob_b}(0.05)$ term cuts on the output of the Level-3 IP-tagger, described in Chapter 3, requiring the probability for the event to have no *b*-jet to be less than 0.05.

background sample	number of events
$b\bar{b}j$ exclusive	900008
$b\bar{b}jj$	948322
$b\bar{b}b\bar{b}$	2210329
$c\bar{c}j$ exclusive	182745
$c\bar{c}jj$	179893
$b\bar{b}c\bar{c}$	453239
$t\bar{t}$	120107
$Zb\bar{b}$	89930

Table 5.2: The number of events simulated for each background process.

to account for multiple interactions. The number of events simulated for each of the background and signal samples is summarized in Tables 5.2 and 5.3.

5.4.1 Corrections Applied to MC

The MC samples were corrected using the standard DØ algorithms [38, 74] to account for the differences in efficiency, resolution and b-tagging between data and MC. The standard DØ algorithms for reproducing the jet energy resolution and jet finding efficiency were applied to the MC. Trigger turn-ons were derived on data for the jet and b-tagging triggers and applied to MC. Taggable jets in MC were reweighted to reproduce the p_T and η distribution measured in data, as described in Section 5.4.1.1. For full details see [22].

The TRFs derived for the NN b-tagger described in the previous Chapter are used to tag MC with the correct efficiency. A comparison was made between this method and direct tagging with the scale factor and the two approaches were found to be comparable. The TRF method is preferred since it preserves the full statistics of the MC samples.

5.4.1.1 Taggability

Before trying to b-tag a jet, the jet is required to be taggable. The taggability criteria demand that the calorimeter jet be matched to a track-jet ($\Delta R < 0.5$). The track-jets are built with ‘good tracks’: they must have $p_T > 0.5$ GeV (one of them $p_T > 1$ GeV), hits in the SMT detector and their impact parameters along the beam axis, z, and in the x-y plane are required to be smaller than 0.2 cm and 0.4 cm, respectively. The taggability in MC must be corrected to match that in

m_h (GeV/c^2)	number of events
90	244059
100	267678
110	246503
115	265917
120	266919
130	268473
140	254434
150	258255
160	267789
170	265205
180	279389
190	267014
200	295527
210	266620
220	319234
260	269987

Table 5.3: The number of events simulated for each tested signal mass.

data; approximately 85% of jets are taggable in data and around 90% in MC. The taggability is measured in data in bins of jet p_T and η . The measured taggability distributions are shown, with the overall mean efficiency factorised out, in Figure 5.2. The taggability factor obtained from the ratio of data and MC taggabilities is used to randomly decide if a given taggable MC jet is to be considered taggable in the analysis, according to its p_T and η.

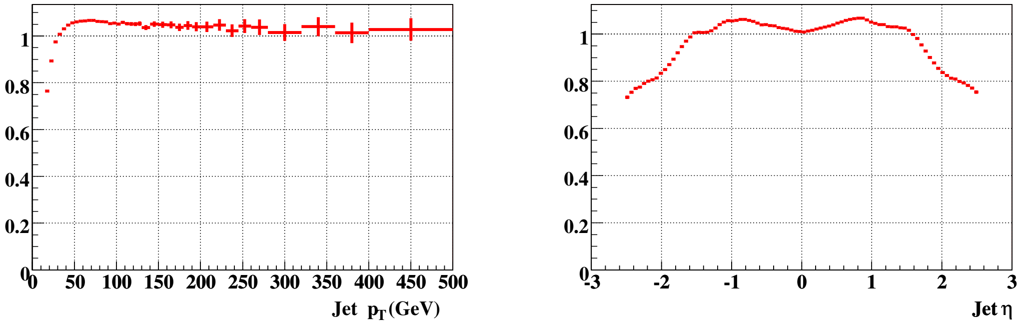

Figure 5.2: The data/MC taggability correction factors parametrised as a function of jet p_T (left) and jet η (right).

The overall taggability correction factor, F, is given by:

$$F = F(p_T) \times F(\eta) \times \frac{85\%}{90\%} \tag{5.1}$$

Where $F(p_T)$ and $F(\eta)$ are the normalised correction factors parametrised in p_T and η as shown in Figure 5.2, and the final term is the overall correction factor.

5.4.2 Signal Simulation

As MSSM and SM Higgs boson production processes are similar, the Pythia SM process $gb \rightarrow hb$ is used. The Pythia simulation also includes a good simulation of ISR and FSR. After including ISR and FSR the two processes, $gb \rightarrow hb$ and $gbb \rightarrow hbb$, are indistinguishable and so the former is used to produce events which are normalised to the NLO cross-section including both processes calculated from MCFM [75]. The kinematic distributions for the Pythia events are also reweighted, in η and p_T, to reproduce those from MCFM.

Though the width of the Higgs is larger in the MSSM than the SM for most of the parameter space it is less than the experimental resolution and so will have limited effect. At large $\tan\beta$ the width becomes non-negligible and so this effect is included in the submitted result.

5.5 Background Modelling

An accurate description of the background is essential to any search. The following sections describe the background composition, overall background normalisation and the background shape for the submitted analysis.

5.5.1 Background Composition

To establish the background composition the MC background H_T distribution is fit to the data in the three-jet channel. The H_T of an event is simply the scalar sum of the constituent jet p_Ts. The fit is performed for samples with between one and three b-tags, for multiple b-tagger operating points.

The background components involving c-quarks are closely related to those with b-quarks and differences due to the differing quark mass become negligible for $H_T >$ 50 GeV. As a result the unknown fractions of cjj and ccj[2] can be fixed relative to the bjj and bbj fractions respectively and the bcc and bbc can be fixed relative to the bbb sample. The ccc sample will be present at the same level as the bbb, which is only significant in the 3-tag sample. Each c-quark will be a factor ∼3-10 less likely to be tagged than a b-quark and so the 3-tagged ccc will be no larger than 3% of the 3-tagged bbb sample. Since 3% is within the errors of the fit process the ccc component is neglected. With these assumptions three variables remain: x_{bjj}, x_{bbj} and x_{bbb}, where x_{bkg} is the fraction of the multi-jet background made up of the given component. The background fractions must sum to unity:

$$\sum_{bkg} x_{bkg} = 1 \tag{5.2}$$

Measuring the number of events in data for zero to three b-tags allows the construction of three further simultaneous equations for each b-tag operating point of the form:

$$\sum_{bkg} x_{bkg} \times \epsilon_{nTags}^{btagOP}[bkg] = N_{nTags}^{btagOP}/N_{tot} \tag{5.3}$$

where $\epsilon_{nTags}^{btagOP}[bkg]$ is the efficiency to tag the given background with $nTags$ at the $btagOP$ operating point. As we use nine operating points the system appears to be highly over-constrained having three unknowns and $1 + 3 \times 9 = 28$ constraints. In fact the different b-tagging operating points are clearly correlated and it is also

---

[2] Here cjj indicates the background component containing one c-jet and two light jets, u, d, s or gluons, in any p_T order. ccj, bjj etc. are to be interpreted similarly.

only possible to constrain the level of bbb in the 3-tag sample. The fractions are determined by a chi-squared minimisation method using the full covariance matrix to account for these correlations. Figure 5.3 shows the result of the sample composition fit for between zero and three *b*-tags for the Tight NN operating point.

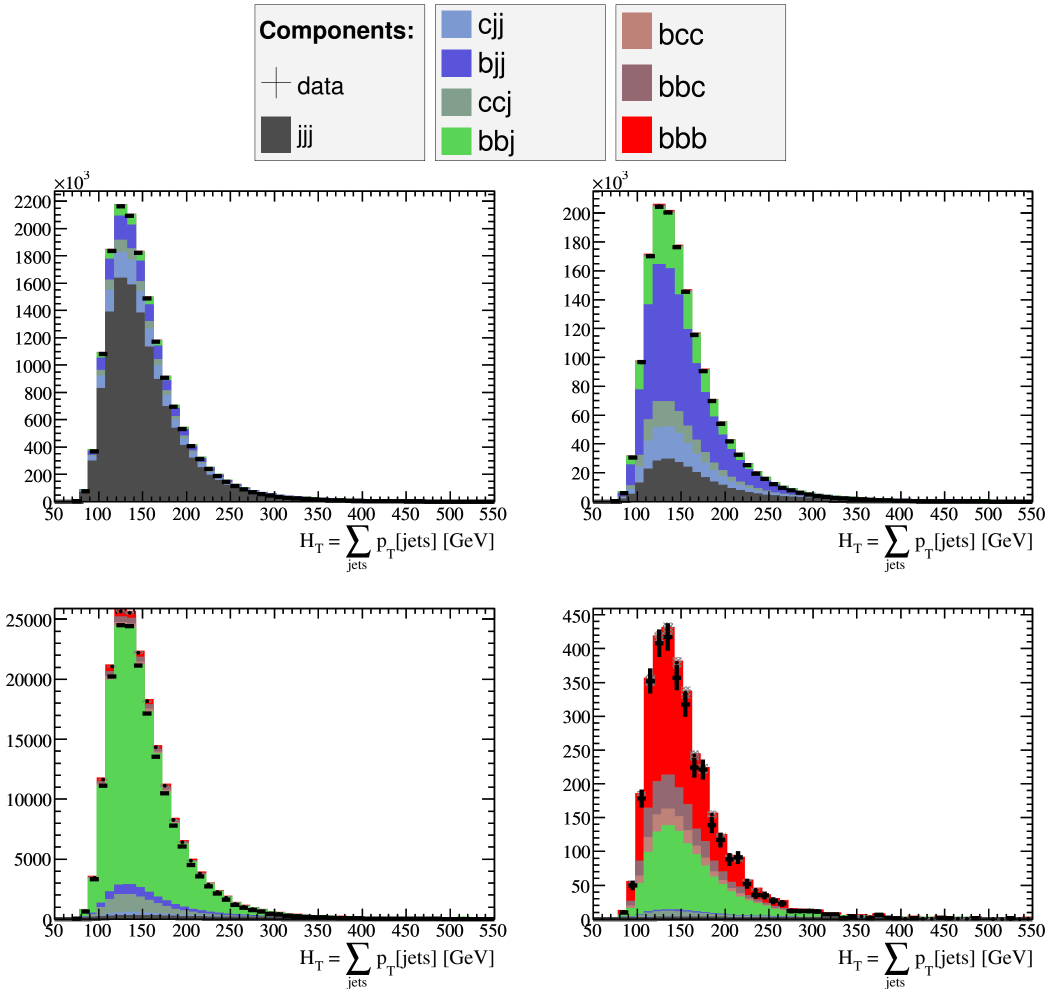

Figure 5.3: The result of the background composition fit for the Tight NN operating point. The component plots refer to events with no tag requirement (top left), at least one (top right), two (bottom left) and three (bottom right) *b*-tags.

The result of the study is used to scale the different background components relative to each other.

5.5.2 Background Level

In the previous iterations of the analysis the background level in MC was determined by fitting the MC to the data in a signal-free region. With the increased power of the NN *b*-tagger the composition of the 3-tag sample has changed and no signal-free region with enough statistics to allow such a fit remains. As a signal-free region with sufficient statistics can no longer be determined the absolute level of the background remains unfixed and only the shape is used to set the limit.

5.5.3 Background Shape

The ratio of the 3-tag and 2-tag mass distributions is referred to as the 3/2-tag ratio. The 3/2-tag ratio derived from background MC can be used in the same way as the 2 → 3-tag probability in the preliminary result, to transform the data 2-tag mass shape into an expected 3-tag shape:

$$Shape^{exp}_{3-tag}(m_h, D) = \frac{Shape^{MC}_{3-tag}(m_h, D)}{Shape^{MC}_{2-tag}(m_h, D)} \times Shape^{data}_{2-tag}(m_h, D) \tag{5.4}$$

where the *Shape* functions, are parametrised in terms of Higgs mass m_h and the multivariate discriminant D, to be discussed later. Integrating over the discriminant variable gives the 3/2-ratio for MC in the 'high-discriminant' region:

$$ratio^{MC}_{3/2-tag}(m_h) = \frac{\int^{D=1}_{D=D_{cut}} Shape^{MC}_{3-tag}(m_h, D)}{\int^{D=1}_{D=D_{cut}} Shape^{MC}_{2-tag}(m_h, D)} \tag{5.5}$$

The value D_{cut} is the lower bound of the high-discriminant region and is optimised depending on the discriminant used (likelihood or NN). In the absence of a signal the ratio is expected to have a linear gradient. To reduce the effect of statistical fluctuations the ratio is fit with a polynomial and this fit function is used as the 2 → 3-tag probability. This ratio is independent of the unknown normalisation of the background. In addition some sources of systematic error will be removed as their effects cancel out in the ratio. A small excess in the 3-tag distribution will appear as an enlarged deviation from the expected linear gradient in the ratio.

Finally the shape of the expected 3-tag invariant mass distribution in the high-discriminant region, $Shape^{exp}_{3-tag}(m_h)$, is given by:

$$Shape^{exp}_{3-tag}(m_h) = ratio^{MCfit}_{3/2-tag}(m_h) \times Shape^{data}_{2-tag}(m_h) \tag{5.6}$$

This expected shape is used as the background-only hypothesis in the limit setting.

5.6 Event Selection Optimisation

The following sections describe how the jet selection, *b*-tagging operating point and number of candidate Higgs pairings per event were optimised. The likelihood discriminant, the multivariate technique used in the submitted result, is also introduced.

5.6.1 Jet Selection

The initial skim criteria $p_T > 15$ GeV originates from the minimum requirement for b-tagging. Since the Higgs is a relatively heavy object at least two jets from signal events are expected to have a harder p_T spectrum than background jets. Increasing the p_T cut for all jets above the minimum 15 GeV, and increasing it further for the lead jet pair, reduces the QCD background. Several jet cut scenarios were considered for the submitted result, the final selection for the likelihood branch of the analysis requires all jets to have $p_T > 20$ GeV and the lead pair to have $p_T > 25$ GeV. The same set of scenarios was considered during the optimisation of the NN analysis branch and the selection used is discussed in Section 5.7.2.5.

5.6.2 b-tagging Operating Point

5.6.2.1 Introduction

The NN b-tagger has a series of operating points with different efficiency and fake rate (see preceding Chapter). In order to find the optimal operating point for this analysis studies were performed for both ICHEP 06 and the final publication. Initially a simple study was performed which preceded the background composition study described above. For this first study the background was estimated from data. Subsequently a second set of studies was carried out using the background model derived in Section 5.5.1. For both studies the significance of the mass peak due to a fixed size injected MC signal is used to find the preferred working point. As the simulation of the b-tag trigger term depends on whether the jet satisfies the offline b-tagging criteria this trigger parametrisation was rederived for each offline working point tested.

For the purpose of these studies, which were performed before the final analysis technique was determined, significance was defined simply as $S/(\sqrt{B} \oplus \epsilon B)$, where S and B are are the number of signal and background events in the 3-tag sample in a $\pm 1.5\sigma$ mass window around the signal, with the mean mass and width derived from a fit to the signal MC mass spectrum. The systematic error, ϵ, used was a fixed value of 3% which was the best estimate at the time. For the initial study the highest p_T jet pair (the 'lead' pair) was used. All available Higgs candidate pairs from each signal and background event were used for the full MC study. The MC study gives a fast and early indication of the expected best working point and allows more complicated b-tagging criteria to be investigated. The study using a background estimate from data is interesting for comparison since it is not affected by data/MC differences

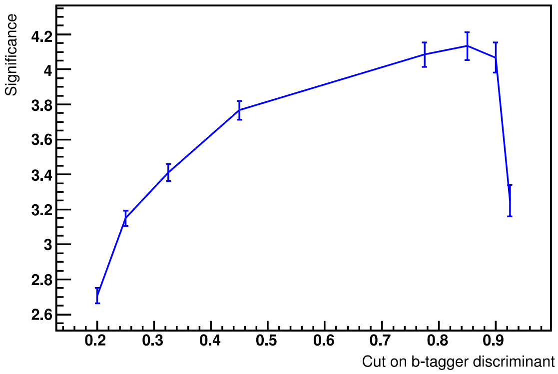

Figure 5.4: The significance seen with an injected MC signal for the ICHEP 06 analysis as a function of the NN operating points. Here the background used in the significance calculation is taken directly from data.

5.6.2.2 Optimised b-tagger Operating Point Using Background from Data

Before the introduction of a multivariate discriminant the signal fraction in the final 3-tag sample is expected to be small. At this stage the data distribution is the best estimate of background and can be used with a large MC signal to form a $S/\sqrt{B}$ significance, as described above. Since $\tan\beta$ can be factorised out of the significance we can choose a large value of $\tan\beta$ and inject a MC signal large enough to make any real signal present in data negligible. The shape of the significance curve is in any case independent of the absolute scale. This simple method gives a preferred b-tagger operating point which is independent of the uncertainties on the background composition and can be derived before the final analysis technique is determined. Figure 5.4 shows the example significance produced for a given MC signal injection against the cut applied to the NN b-tagger output. The example shown is for a 140GeV Higgs and $\tan\beta = 70$. Several masses (between 90 and 160 GeV) were investigated and the choice of operating point was found to be independent of the signal mass. The study indicates that, as expected, tighter operating points are preferred. The tightest available operating point is not recommended, the next three tightest are within errors equivalent and give the best significance.

5.6.2.3 Optimised b-tagger Operating Point Using the Full MC Background Composition

The second phase of the operating point optimisation attempts to achieve two things, firstly to reduce the errors in order to be able to differentiate between the results from the three tight working points suggested above and secondly to investigate the possibility of improvements from more complicated tagging procedures such as a combined event tag. This part of the study was conducted after the optimisation of the jet p_T cuts for the final analysis. With these cuts the signal is predominantly in the 3-jet channel and therefore the study was confined to that channel. Taking the background distribution from MC increases the statistics considerably. Several combinations of tags were also considered to investigate if combining the tag information from several jets could enhance the background rejection. The following alternative scenarios where tried, where T_n is the NN b-tagger output for the n^{th} jet:

- **Scenario A** cut on T_1, T_2 and T_3 with the same operating point (as in the previous section)
- **Scenario B** cut on T_1, T_2 and T_3 with different operating points
- **Scenario C** cut on $log(T_1 + T_2)/log2$ and T_3
- **Scenario D** cut on $log(T_1 + T_2 + T_3)/log3$

The jet order must also be specified, ordering with jet p_T or with NN b-tagger output would give different results for the second and third scenarios above; both options were investigated. Figure 5.5 shows the significance as a function of the b-tag cut applied for scenarios A and D, the two best performing scenarios. None of the combinatorial event tags tried gave a better significance. A very tight cut on the sum of the NN output values for the three jets can, within errors, equal the best significance from tagging the three jets independently, but the second differential at the optimum point is larger for the combined tag meaning the optimum point is less stable and so discriminating with this variable is less reliable. The composition of the background is fixed from the study in Section 5.5. However the overall background is not normalised to data since no signal free region is available, as a result the background level is low by a constant factor which results in the higher significances in Figure 5.5 compared to Figure 5.4.

Since there are significant backgrounds for all levels of b-content it is understandable that it proves impossible to gain by loosening any of the tags. Any of the cuts on a linear combination of tag values result in a limited loosening of the criteria since one particularly tight tag can allow an event with a loosely tagged jet to pass. Light jets are an order of magnitude more common in the QCD backgrounds than b-jets and so high rejection operating points, 'tight' b-tags, are needed to remove the overwhelming jjj and bjj backgrounds. Once the bbj background component ceases to be dominant, i.e. once three 'tight' b-tags have been required, the bbb component dominates. It is clearly impossible to reduce the bbb background with respect to

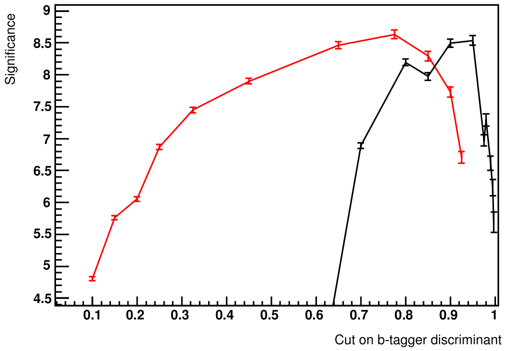

Figure 5.5: The significance seen with an injected MC signal and background from the background composition study. Significance is shown after three b-tags at one operating point (red) and after a cut on the sum of the b-tagger output for the three jets (black).

the signal by tightening the b-tagging since signal and background have identical b-content. At this point further tightening of the working point simply reduces the statistics for both signal and background.

5.6.2.4 Conclusion of b-tagger Operating Point Studies

The optimal working point for the NN b-tagger for this analysis was found to be Tight, applied to all three b-tagged jets. Alternative event tags using the combined information from multiple jets were also investigated but due to the large QCD backgrounds at all levels of b-content these were not preferred to requiring tags with a uniform operating point on each individual jet. The study was repeated with the final full analysis using a reduced number of operating points and the same conclusion was drawn [22].

5.6.3 Number of Higgs Candidate Pairs

For a signal event in data the jet pair coming from the Higgs decay is not known. MC suggests that in a significant fraction of events ($\sim 30\%$) the Higgs pair is not the leading p_T pair. In order to recover these signal events additional Higgs candidates formed from other jet pairs can be included. Three scenarios were considered:

One pair - use the lead p_T pair.

Two pairs - use the lead pair and the pair made from the highest and third highest p_T jets.

Three pairs - use both pairs given above and the pair made from the second and third highest p_T jets.

Since the lead pair accounts for the majority of true Higgs pairs in signal MC only combinations including the lead pair need be seriously considered. Higgs candidates for signal and background MC constructed using each of the three sets of pairs above were run through the full analysis chain including the likelihood discriminant and used to set limits. The best expected limits were determined using two candidate pairs per event [22].

5.6.4 The Likelihood Discriminant

The main branch of the final analysis uses a likelihood discriminant as the multivariate technique in the final event selection. The alternative NN branch is discussed in Section 5.7. The likelihood ratio, R_{LH}, is computed using Equation 5.7 with probability density functions (PDFs) generated by fits to the variable distributions in MC:

$$R_{LH}(x_1, x_2...x_n) = \prod_{n}^{i=1} \frac{P_i^{sig}(x_i)}{P_i^{bkg}(x_i)} \tag{5.7}$$

where x_i are the n input variables and P_i are the PDFs for signal and background. The ratio is converted to a discriminant variable, ${Y_{LH}}^3$, designed to peak at one for signal and zero for background:

$$Y_{LH} = \frac{R_{LH}}{1 + R_{LH}} \tag{5.8}$$

The likelihood uses six kinematic variables, as listed below:

$\Delta\eta_{b_1 b_2}$ - the separation in η of the Higgs candidate jets.

$\Delta\phi_{b_1 b_2}$ - the separation in ϕ of the Higgs candidate jets.

$|\arccos(\vec{b_1} \cdot \vec{h})|$ - the angle between the leading Higgs jet and the Higgs candidate.

$|p_{b_1} - p_{b_2}|/|p_{b_1} + p_{b_2}|$ - the momentum balance of the jets forming the Higgs candidate.

Y - the rapidity of the Higgs candidate.

---

[3] This replaces the generic dicriminant, D, in Equations 5.4 and 5.5

Sphericity - the sphericity of the event.

All were well reproduced in MC. Comparisons of the variables in data and MC are shown in Section 5.7.1. A full description of the likelihood discriminant can be found in [22].

5.7 Developing the Neural Network Event Selection

The NN selection was developed in parallel with the likelihood selection. Having two multivariate methods available to produce a final selection helps to confirm that the result we see is not an artifact of the implementation of one of the techniques. Neural networks consist of multiple nodes processing linear combinations of input variables simultaneously. In theory a NN method is more likely to respond to subtle correlations between input variables than the likelihood discriminant which considers each variable independently. It may therefore be possible to increase the background rejection by moving to a NN discriminant. The more intuitive likelihood would then become the cross-check method. With the aim of interchangeable discriminants in mind the preselection scenarios (p_T cuts, Higgs pairings per event) under consideration for the submitted analysis were investigated for the NN, allowing the NN to slot into the main analysis. In addition the optimal jet selection and number of Higgs candidate pairs for the NN analysis was also found in parallel with the NN configuration optimisation. As in the preceding Chapter the NN was constructed and optimised using the `TMultiLayerPerceptron` package provided in `ROOT` [65].

5.7.1 Description of Variables

The following kinematic variables were considered for inclusion in the NN. The jets b_1 and b_2 are those used to reconstruct the Higgs candidate, in p_T order. b_3 is the remaining jet not used to form the Higgs candidate. The variables used in the likelihood are marked with an asterisk.

$\Delta\eta_{b_1 b_2}$ - the separation in η of the Higgs candidate jets. $\star$

$\Delta\phi_{b_1 b_2}$ - the separation in ϕ of the Higgs candidate jets. $\star$

$\arccos(\vec{b_1} \cdot \vec{h})$ - the angle between the leading Higgs jet and the Higgs candidate. $\star$

$|p_{b_1} - p_{b_2}|/|p_{b_1} + p_{b_2}|$ - the momentum balance of the jets in the Higgs candidate. $\star$

Y - the rapidity of the Higgs candidate. $\star$

$\vec{b_3} \cdot \vec{h}$ - cosine of the angle between the third jet and the Higgs candidate.

2^{nd} Fox Wolfram Moment, H_2 - a quantity relating to the spherical symmetry of the jets in the event, defined in Equation 5.9:

$$H_2 = \sum_{ij} \frac{|p_i||p_j|}{E_{vis}^2} \frac{1}{2} \left(3\cos^2\theta_{ij} - 1\right) \tag{5.9}$$

where E_{vis} is the total visible energy in the event, θ_{ij} is the angle between the jet directions and the sum over (i, j) includes all jet pairs.

Sphericity - the sphericity of the event. $\star$

Figure 5.6 shows the variable distributions for data compared to signal and background MC samples, where three b-tags have been required. The signal MC is for a 190 GeV Higgs. For the signal MC only the correct pairing of Higgs jets is shown, for data and background MC all pairs are shown. The area under each histogram is normalised to unity. The same distributions are shown in Figure 5.7 for candidate pairs within a $\pm 1.5\sigma$ mass window around the mean of a Gaussian fit to the MC signal pair mass distribution. Good agreement can be seen between data and background MC for all variables in both regions. It is also clear that individually each variable can offer only limited discrimination. Some disagreement is seen in the comparison of $\Delta\eta_{b1b2}$ but the data look less signal-like than the background MC and so the discriminating power measured in MC will be conservative.

5.7.2 NN Optimisation

The configuration of the NN and the preselection of candidate pairs to be considered were optimised simultaneously. The following sections give snapshots comparing NNs produced whilst varying individual components of the configuration, these are not the complete set of NNs produced but give an idea of all the comparisons which were made. In the following sections the plots refer to the 3-jet channel unless otherwise noted, using a NN in the 4/5-jet channel was also investigated but the signal is considerably smaller.

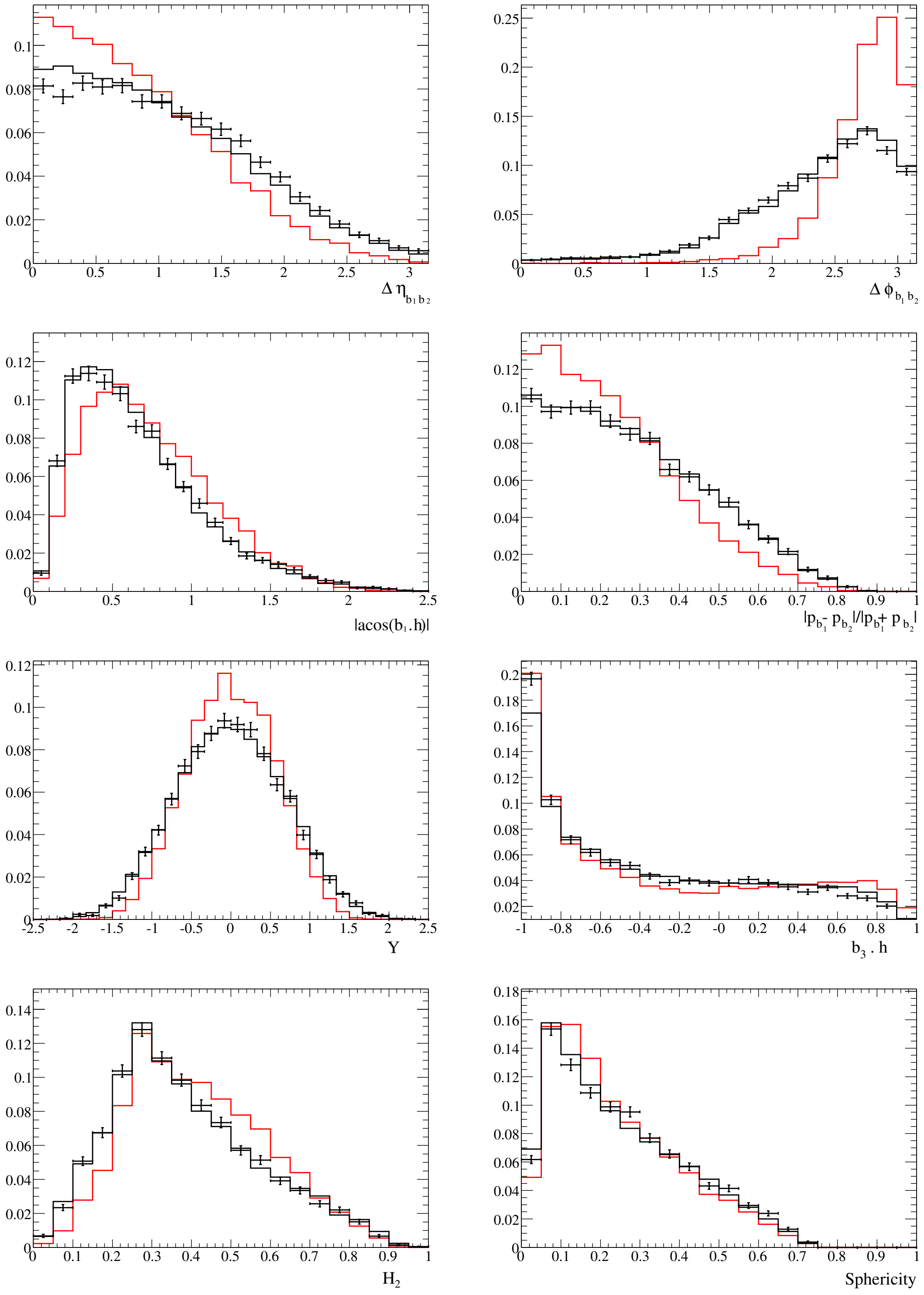

Figure 5.6: Input variable distributions for all jets for data (black points), background MC (black histogram) and correct pairs only in signal MC (red histogram).

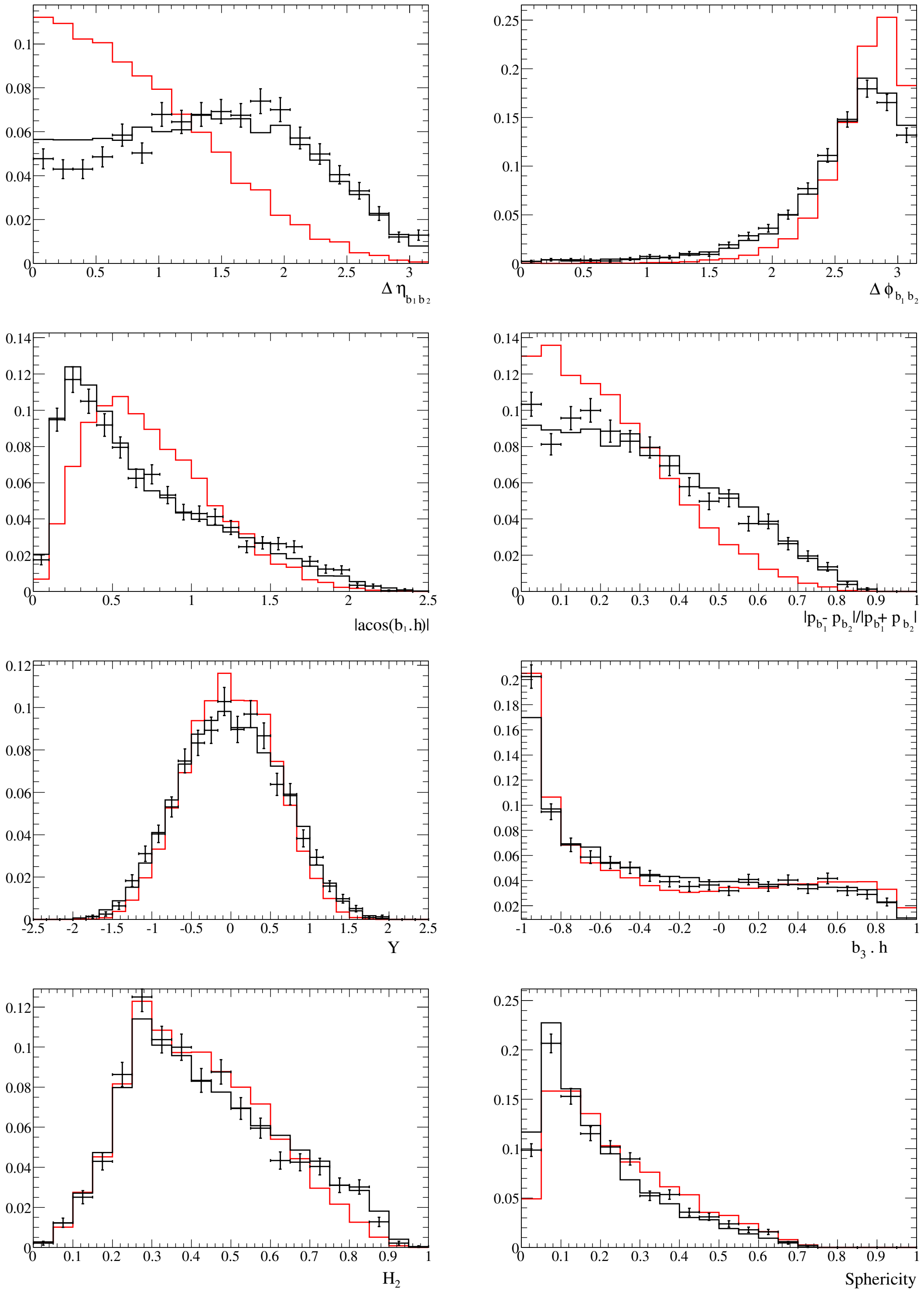

Figure 5.7: Input variable distributions for jets within the signal mass window for data (black points), background MC (black histogram) and correct pairs only in signal MC (red histogram).

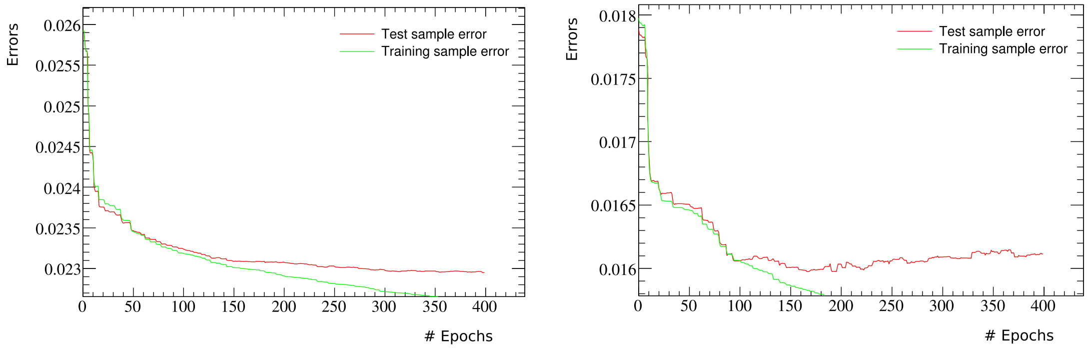

Figure 5.8: Example plots of the training sample and test sample errors against training epoch are shown for a NN trained in the 3-jet (left) and 4/5-jet channels (right). The errors are the fraction of incorrect NN outputs at the current epoch given an NN cut of 0.5.

5.7.2.1 Training Method and Number of Epochs

The BFGS training method [67] was again used since it is considered to be the most efficient training algorithm (see preceding Chapter). A neural network was independently trained for each preselection configuration, for each variable selection and for the 3-jet and 4/5 jet channels. The networks were trained using a MC sample consisting of the signal at a given mass and a background composed of the three most significant MC samples (bbj, bbjj and bbb) with relative weights as determined in Section 5.5. The total number of background candidate pairs was weighted to be equal to the number of signal pairs at the specific training mass. Only the correct Higgs pairing was used for signal events in the training sample. Half of the MC available was kept in reserve and used to test the NN, this MC is referred to as the test sample. The NN was considered trained when the fraction of incorrect outcomes (for a cut at 0.5) on the test sample no longer decreased with further training. Figure 5.8 shows the error, the fraction of incorrect NN outputs, for the test and training samples for one representative Higgs mass. The error for the training sample continues to decrease as the network becomes over trained but the error on the test sample reaches a minimum. The network training is recorded for each epoch and the network with the minimum test sample error is taken as the final network. Four hundred training epochs was found to be sufficient to always find a minimum in the test sample error.

5.7.2.2 Optimisation Benchmark

Curves of signal efficiency against background pass-rate were produced for each NN set-up and the likelihood method by considering cuts on the discriminant variable over the complete range. The signal and background samples are MC, where the background composition is specified by the study in Section 5.5. For a given background rate a higher signal efficiency will produce a greater final significance and so curves toward the top-left in the following performance plots are preferred. For the final optimisation the point which produces the highest significance on each curve is taken and compared. The chosen cut separates the candidate Higgs pairings into a low- and a high-discriminant region. To calculate the significance the low-discriminant region, which is signal free, is used to produce a 2 → 3-tag background-only tag rate function (TRF). This tag rate function is applied to the high-discriminant 2-tag distribution to produce a prediction for the background-only high-discriminant 3-tag distribution. This prediction can be subtracted from the actual high-discriminant 3-tag distribution to get the observed signal. A simple signal significance is calculated as in Section 5.6.2:

$$
\begin{aligned}
S + B &= N_3^{high} && (5.10)\\
\text{predicted } B &= N_2^{high} \times \frac{N_3^{low}}{N_2^{low}} && (5.11)\\
\text{significance} &= \frac{S}{\sqrt{B} \oplus 3\% \cdot B} && (5.12)\\
&= \frac{\mathrm{N}_3^{high} - \left(\mathrm{N}_2^{high} \times \frac{\mathrm{N}_3^{low}}{\mathrm{N}_2^{low}}\right)}{\sqrt{\mathrm{N}_2^{high} \times \frac{\mathrm{N}_3^{low}}{\mathrm{N}_2^{low}} + \left(\epsilon \times \mathrm{N}_2^{high} \times \frac{\mathrm{N}_3^{low}}{\mathrm{N}_2^{low}}\right)^2}} && (5.13)
\end{aligned}
$$

Where $\mathrm{N}_{2/3}^{high/low}$ is the number of invariant mass pairs counted in the high- or low-discriminant, 2- or 3-tag distribution. The systematic error on the background, ϵ was estimated at the time to be $\sim 3\%$. The number of pairings are counted in a $\pm 1.5\sigma$ window generated from a Gaussian fit to the mass distribution for the correct Higgs pairing in the signal MC. The signal cross-section is normalised using $\tan\beta = 160$. This value is larger than the existing limit but is used as a 'best case' to show what the 3/2-ratio distributions would look like in the case of a large signal. As mentioned previously the $\tan^2\beta$ signal enhancement in the $S/\sqrt{B}$ significance can be factorised out and so the choice of $\tan\beta$ does not effect the optimisation of the discriminant.

5.7.2.3 NN Optimum Working Point

The working point which produces the highest significance was found by calculating the significance described above while stepping through cut values on the NN output between -0.1 and 1.1 in steps of 0.05. The resulting curve of significance against NN output cut can be seen, along with the actual NN output distributions for signal and background, for the final optimised NN set-up (as an example) in Figure 5.9. In this case the optimum cut would be ∼ 0.55.

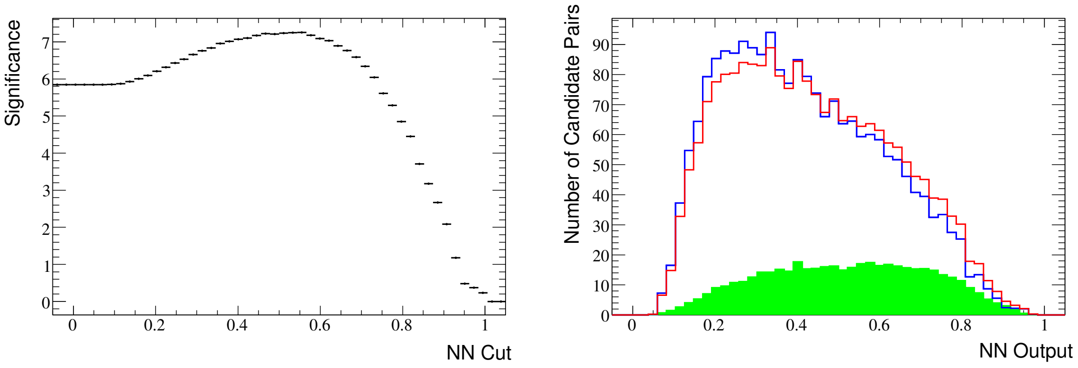

Figure 5.9: The significance calculated with a given MC signal for the full range of cut values (left) and the actual output of the NN (right) where the signal-plus-background (red) is normalised to the same level as the background-only (blue) shape. The signal-only shape is shown (green) for the given signal size. This example shows the distributions for the final optimised NN set-up.

5.7.2.4 NN Structure

The number of hidden nodes should be greater than the number of input nodes, increasing the number of hidden nodes increases the CPU time required to train and to run the NN. The number of hidden nodes was varied between 7 and 28. This optimisation was performed with the standard preselection (two jets above 20GeV, one above 25GeV, the leading two jet pairings) in the 3-jet channel. The performance curves for each tested neural net structure are shown in Figure 5.10. It is apparent from the figure that the network structure has little effect on the final power of the network. The network with 21 hidden nodes had the lowest, or joint lowest, background pass rate for any given signal efficiency between 40 and 85% and so this network structure was chosen.

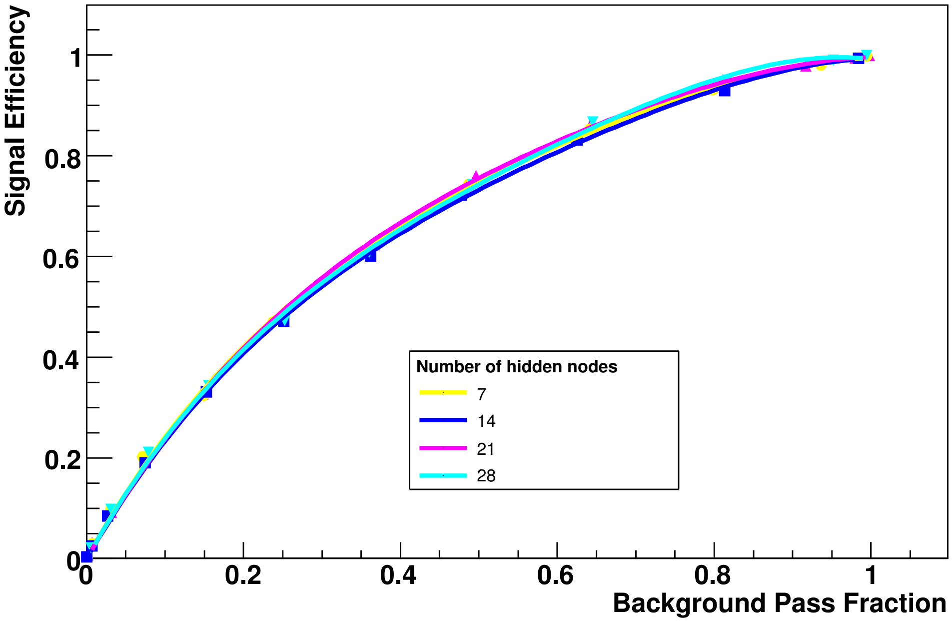

Figure 5.10: Performance curves for a set of NNs produced with different numbers of hidden nodes. The dependence is small and 21 nodes is chosen for the final set-up. Statistical errors are shown but are smaller than the marker size.

5.7.2.5 NN Input Preselection: Jets

Three different jet selection scenarios were considered during the lifetime of the main analysis as given below:

- **Scenario i.** All jets above 15 GeV, the basic requirement for taggability.
- **Scenario ii.** Two b-jets above 25 GeV, all jets above 15 GeV, since the Higgs is a heavy object which should decay to high p_T jets.
- **Scenario iii.** Two b-jets above 25 GeV, all jets above 20 GeV, since the jet from the associated production is also expected to be high p_T.

A NN was specifically trained for each preselection. The 3/2 ratios for MC in the low- and high-NN regions are shown in Figure 5.11 for each set of jet cuts with the corresponding network. The 3/2-ratio for signal-plus-background is calculated including all candidate Higgs pairings for both signal and background since in data we cannot distinguish the correct pair. In the low-NN region the background and signal-plus-background distributions are similar in shape and numerically close. Since the b-content of the background is not very well known we cannot predict the absolute level of the 3/2 ratio and so we would not expect to be able to detect the effect of the signal in this region in data. The different absolute level of the 3/2 ratio for the low-NN region is due to some of the additional combinatorial background from the incorrect pairings from signal events. Whilst coming from signal events this combinatorial background is made up of pairs which do not originate in a Higgs, it is not peaked and has a shape close to background. The low-NN region is as close to a signal free region as we can produce and will be used to confirm that the background in data is behaving as expected. The low-NN region background and signal-plus-background are fit with straight lines. The high-NN region signal-plus-background is fit with a Gaussian plus a straight line with gradient fixed from the background-only fit in the low-NN region. The width of the Gaussian is fixed to be the di-jet energy resolution, which is 28GeV for 90GeV jets.

Table 5.4 shows the values for significance derived from the high-NN region using the low-NN region as a TRF as described in Section 5.7.2.2. Each significance is calculated at the optimum working point for the relevant NN, i.e. it is the best significance calculated for the given jet selection. The third, tightest, jet selection was seen to produce the most significant excess with the given MC signal. This set of jet cuts was chosen for the final analysis and is used for the following optimisations.

jet selection	signal ‘significance’
3 jets above 15 GeV	5.75
2 *b*-jets above 25, 1 jet above 15 GeV	5.76
2 *b*-jets above 25, 1 jet above 20 GeV	5.95
best result with likelihood	7.09

Table 5.4: Significances calculated for three jet selections with a fixed MC signal using the low-NN region to predict the background 2 → 3-tag rate. The error on the significance, mostly due to signal MC statistics, is $\sim 3\% = 0.2$.

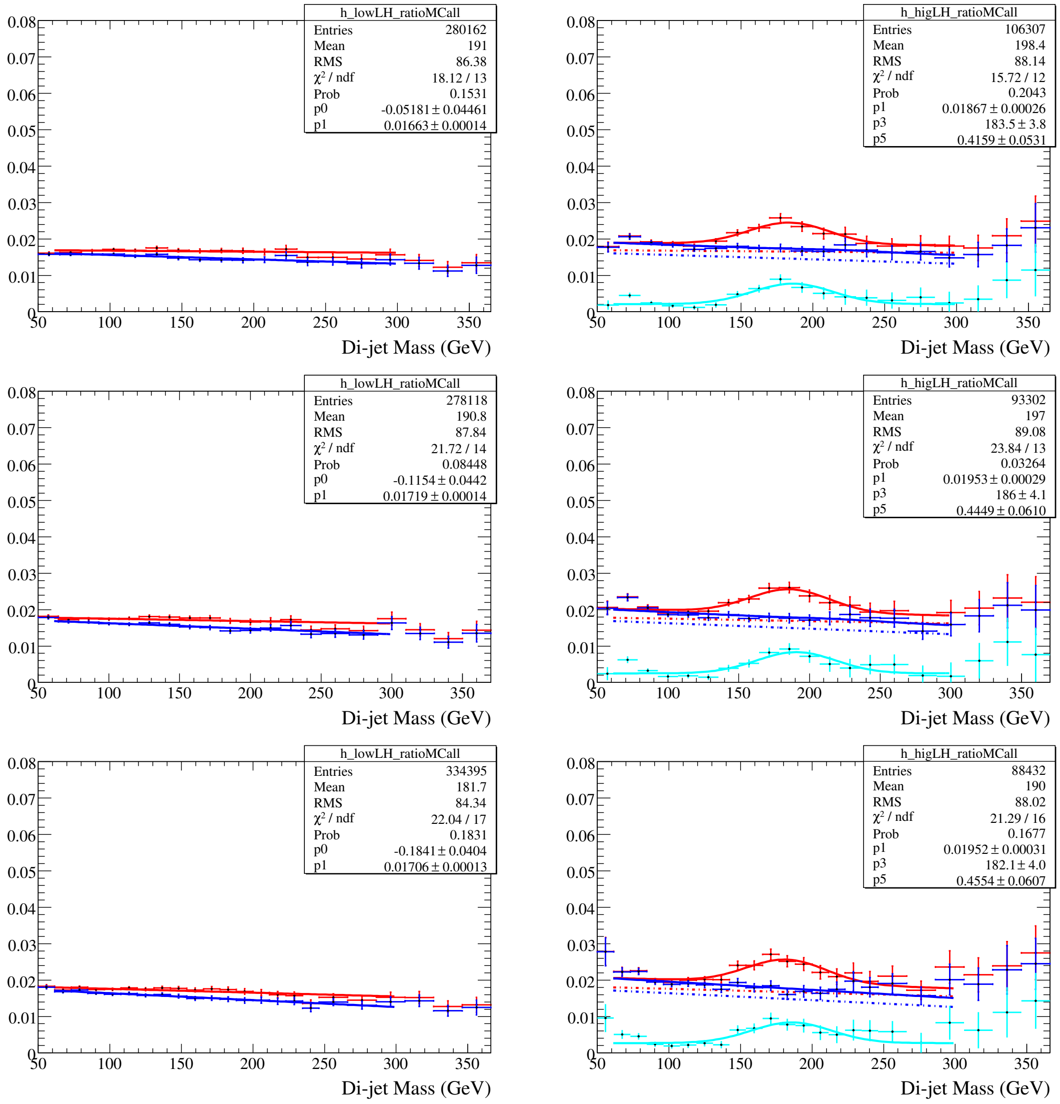

Figure 5.11: The MC 3/2 ratios for the three jet selections (Scenarios i. top, ii. middle, iii. bottom) for events with NN-output < 0.5 (left) and greater than the optimised cut (right). The blue points and fit are for background-only, the red points and fit are for signal-plus-background. The cyan points and fit correspond to signal-plus-background minus background-only.

5.7.2.6 NN Input Preselection: Number of Higgs Pairs

In a 3-jet event there are three possible pairs of jets which could be used to form a Higgs candidate as discussed in Section 5.6.3. To improve the acceptance for correct signal pairings all possible candidate pairs are considered, the final choice of which pairs to keep is optimised using the full analysis chain with an individually retrained NN selection.

Figure 5.12 shows the 3/2-ratios for the low-NN region (output less than 0.5 for the given NN) and for the high-NN region where the cut value is re-optimised for each NN. Again, in calculating the 3/2-ratio, the signal and background MC are treated identically since in data we cannot identify the jets from the real Higgs. The ratios and fits are as explained in the preceding section. The results of the significance calculated by the $2 \rightarrow 3$-tag TRF approach explained above are summarised in Table 5.5. Using only the lead and second pairings is seen to produce the highest significance and is adopted for the remaining studies.

jet selection	signal ‘significance’
Lead pair only	5.52
Lead and second pair	6.19
All pairs	5.95
best result with likelihood	7.09

Table 5.5: Significances calculated for different choices of Higgs pairings with a fixed MC signal using the low-NN region to predict the background $2 \rightarrow 3$-tag rate. The error on the significance, mostly due to signal MC statistics, is $\sim 3\% = 0.2$.

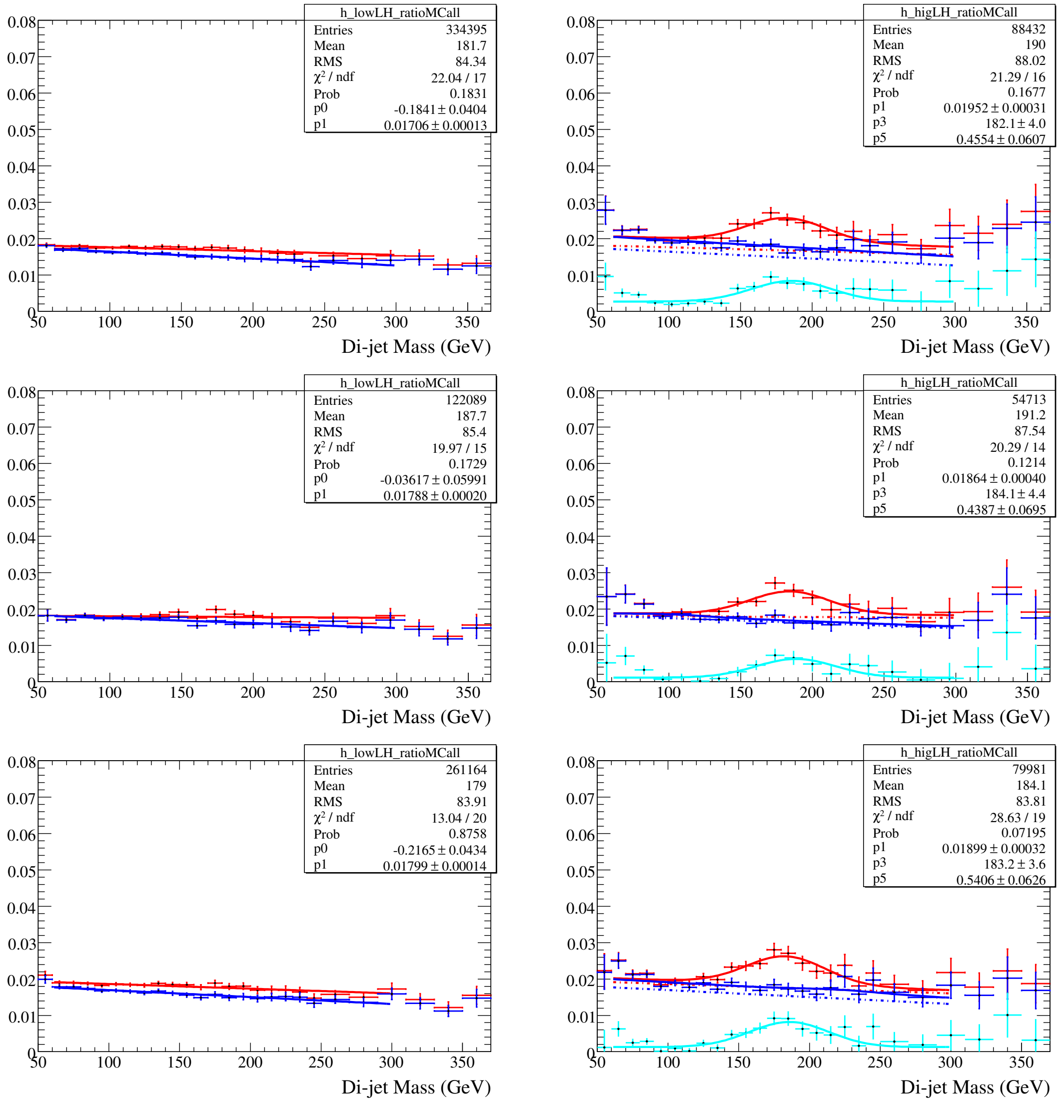

Figure 5.12: The 3/2 ratios for the three choices of Higgs pairings (Lead pair only: top; Lead and second pair: middle; All pairs: bottom) for events with NN-output < 0.5 (left) and $>$ the optimised cut (right). The blue points and fit are for background-only, the red points and fit are for signal-plus-background. The cyan points and fit are for signal-plus-background minus background-only. The jet preselection is as optimised in the previous section.

5.7.2.7 Training Within a Mass Window

The impact of using a mass window during training was also investigated. NNs were trained with and without a mass window on the pairings used in the training. All pairings (including those outside the window) were processed by the resulting networks to produce the complete low- and high-discriminant mass spectra. The window is $\pm 1.5\sigma$ where σ comes from a fit to the MC signal mass distribution. The shape of the background-only distribution is unchanged if a mass window is applied. This suggests that the application of a mass window in the NN training will not generate a false peak in data. Figure 5.13 shows the signal efficiency and background pass fraction for the two network training methods. Training using only candidate pairs within a mass window is seen to be significantly more efficient for a given background rate. The 3/2-tag ratios for the low- and re-optimised high-NN regions are shown in Figure 5.14. In addition to increasing the significance with which the example signal is observed, training with the mass window improves the behaviour of the mass distribution for low masses in the high-discriminant region, improving the quality of the linear fit to the background-only distribution. The toy significance values calculated with the fixed MC signal as described previously are recorded in Table 5.6. As expected the network trained with the mass window gives a better significance.

	signal ‘significance’
No mass window on training	6.19
Mass window for training	7.38
best result with likelihood	7.09

Table 5.6: Significances for NNs produced with and without a mass window on the pairs used for training, calculated with a fixed MC signal using the low-NN region to predict the background $2 \rightarrow 3$-tag rate. The error on the significance, mostly due to signal MC statistics, is $\sim 3\% = 0.2$.

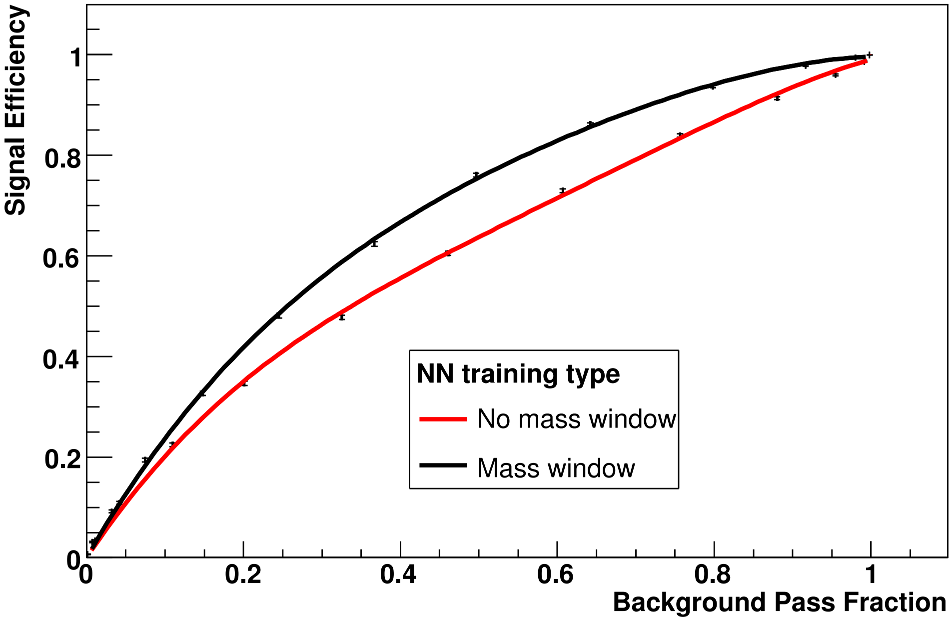

Figure 5.13: Performance curves for NNs produced with and without a mass window on the jet pairs used for training.

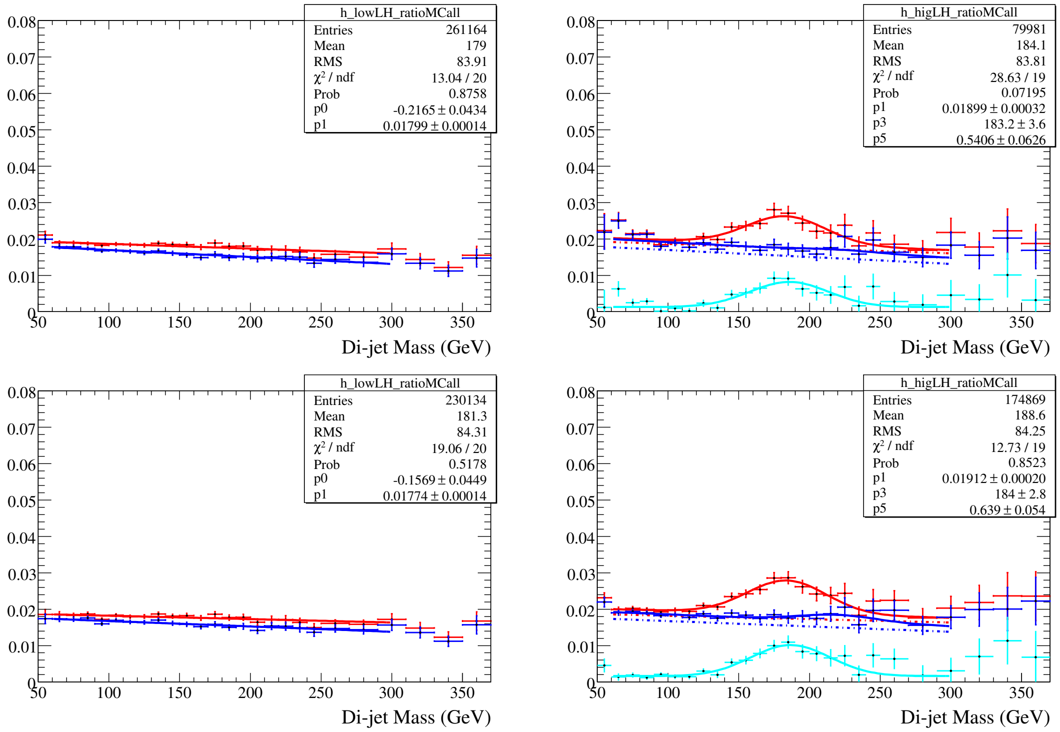

Figure 5.14: The 3/2 ratios with and without a mass window on the Higgs candidate pairs used for training (Without window: top; With window: bottom) for events with NN-output < 0.5 (left) and greater than the optimised cut (right). The blue points and fit are for background-only, the red points and fit are for signal-plus-background. The cyan points and fit are for signal-plus-background minus background-only. The jet and Higgs pair preselection is as optimised in the preceding sections.

5.7.2.8 Variable Selection

The power of variables used in the network is automatically estimated by the `TMultiLayerPerceptron ROOT` package by making small variations to the input distributions and observing the change in the number of incorrect NN outputs [65]. Having used this method to roughly order the variables, the NN is rebuilt and retrained with successive variables removed, weakest first. The performance curves for each network were generated and compared to see which input variables could be dropped from the final network. Such a 'pull down' method is preferred to a 'build up' method when deciding on NN variables since correlations between the variables are a significant source of power for a NN over simpler techniques such as the likelihood. Starting with all available variables ensures that all correlations are included in the initial NN and the reduction in power when a variable is removed includes the loss from correlations. Table 5.7 gives the variables in order, most powerful first, for NNs trained with the final preselection both with and without the mass window on pairs used for training. The application of the mass window changes the order of most powerful variables, however the weakest three remain the same.

Rank	NN with mass window	NN without mass window
1	Higgs Rapidity	$d\phi_{1,2}$
2	$\cos(\theta_{3,H})$	Sphericity
3	$d\phi_{1,2}$	Higgs Rapidity
4	$d\eta_{1,2}$	$\cos(\theta_{3,H})$
5	Sphericity	$d\eta_{1,2}$
6	Momentum balance	2^{nd} Fox Wolfram moment
7	2^{nd} Fox Wolfram moment	$d\theta_{1,H}$
8	$d\theta_{1,H}$	Momentum balance

Table 5.7: The ranking of variables in NNs trained with and without mass windows on training pairs. Increasing rank corresponds to weaker variables.

Figure 5.15 shows the performance curves for NNs trained with the full set of variables and with several reduced sets. Removing the weakest three variables causes only a small reduction in performance. However since the reduction is as large or larger than the error on the performance measurement it was decided to retain all variables for the final network.

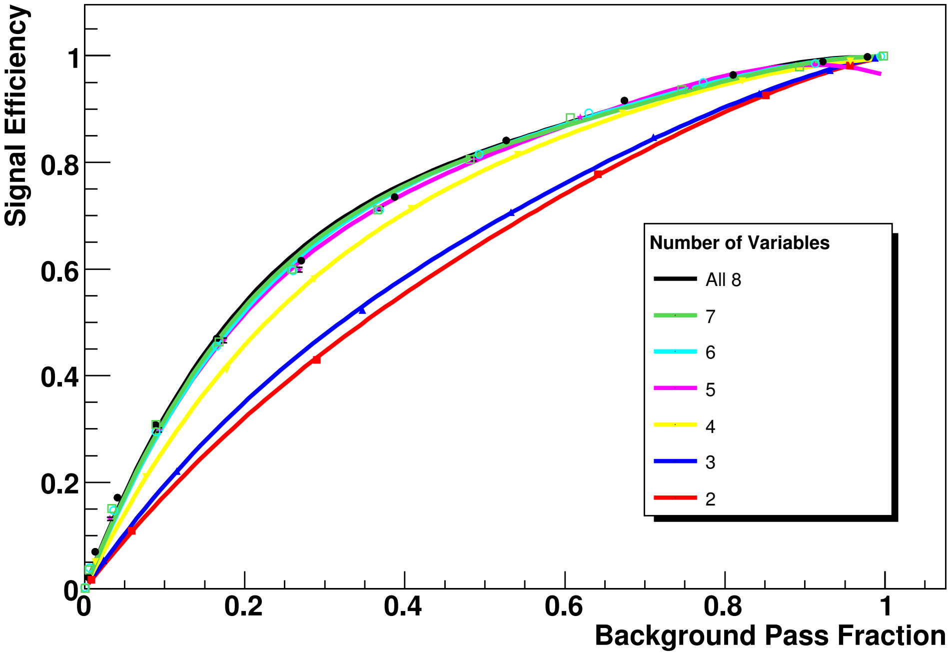

Figure 5.15: Performance curves for NNs trained with the weakest variables progressively removed. Errors are shown but are small. The effect of removing the weakest variable is small, on the same order as the errors. For the final network configuration all eight variables are retained. The selection of pairs used is as optimised in the previous sections.

5.7.3 Data-MC Agreement for the Optimised NN

The NN output for the final best NN configuration is shown in Figure 5.16. Since the absolute level of the background is not well known the distributions for background-only MC and background-plus-signal MC are normalised to the number of events in the data sample. For the background-plus-signal sample the signal is the 190GeV Higgs used to train the network and is normalised with respect to the background component to the correct ratio of NLO cross-sections. The NLO cross-section for the signal component requires a $\tan\beta$ hypothesis and the same value is chosen as is used to calculate the toy significance, $\tan\beta = 160$. This $\tan\beta$ value is above the existing limit up to the highest m_A values investigated and as such the signal-plus-background shape gives an indication of the maximum effect in the NN output distribution due to the presence of a signal. From this plot the distribution of NN output values for data is seen to agree well with both a background-only and background-plus-signal hypothesis. The normalised distribution of signal-only, correct-pair, NN outputs is also shown to indicate the separation power of the network. The distributions are shown for all mass values and for pairs with invariant mass within the $\pm 1.5\sigma$ mass window to show that the agreement is good for both the training sample and the complete analysis sample.

5.7.4 Comparison to the Likelihood

The performance of the NN and likelihood discriminant are compared in Figure 5.17. A small improvement is seen from using the NN which is reflected in the small ($\sim$ 4%) increase in significance seen in Table 5.8. The 3-over-2-tag ratio distributions are shown in Figure 5.18 and the slight increase in the signal can be seen in the high-discriminant region.

	signal ‘significance’
NN	7.38
Likelihood	7.09

Table 5.8: Significances for the optimum NN compared to that from the best likelihood, calculated with a fixed MC signal using the low-NN region to predict the background $2 \rightarrow 3$-tag rate. The error on the significance, mostly due to signal MC statistics, is $\sim 3\% = 0.2$.

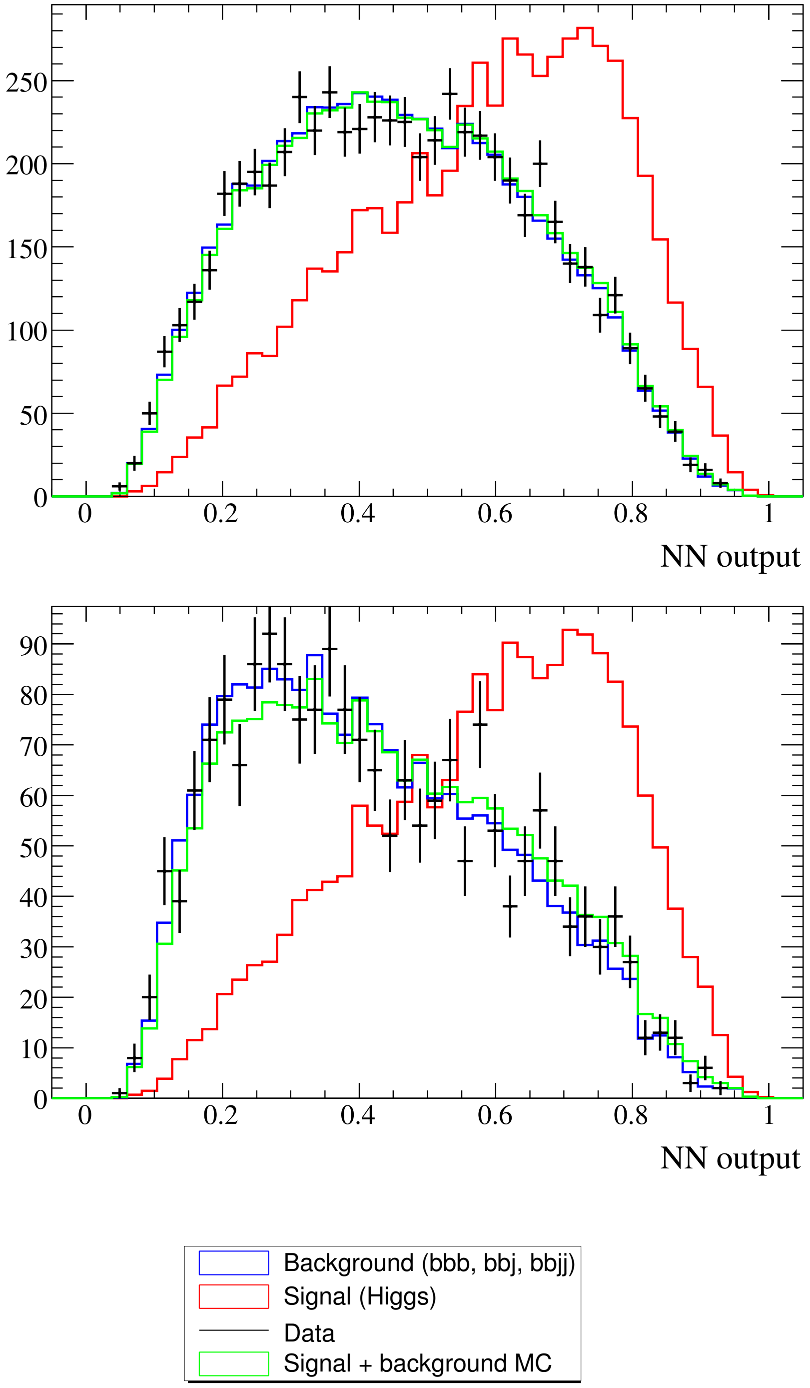

Figure 5.16: The distribution of NN outputs in data (black) compared to the background-only (blue) and signal-plus-background (green) hypotheses. Both hypotheses are normalised to the total number of data events. The signal is a 190 GeV Higgs with $\tan\beta = 160$. The NN output distribution for the correct Higgs di-jet pair in signal events is also shown, normalised to the total size of the data sample, for comparison (red). All candidate pairs regardless of mass are shown in the left-hand plot, while right-hand plot shows only those pairs within the $\pm 1.5\sigma$ mass window used for training. The data is consistent with background-only MC, and with a small injected signal, in both cases. The difference between background-only and signal-plus-background is increased by applying the signal window (lower plot).

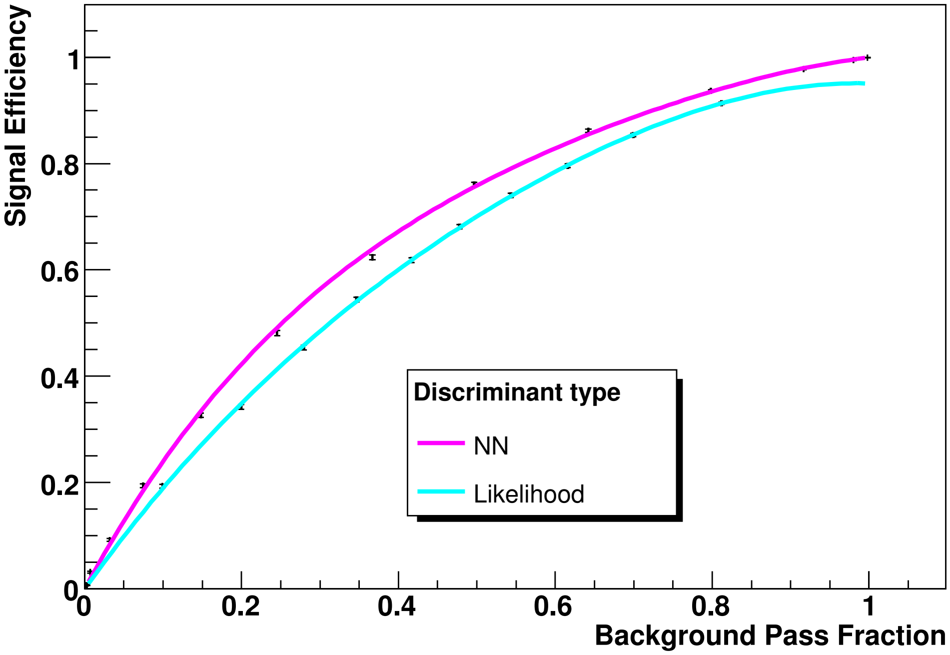

Figure 5.17: Performance curves for the two types of multivariate discriminant. The NN trained with the optimum configuration and preselection is shown in purple compared to the optimum likelihood in cyan.

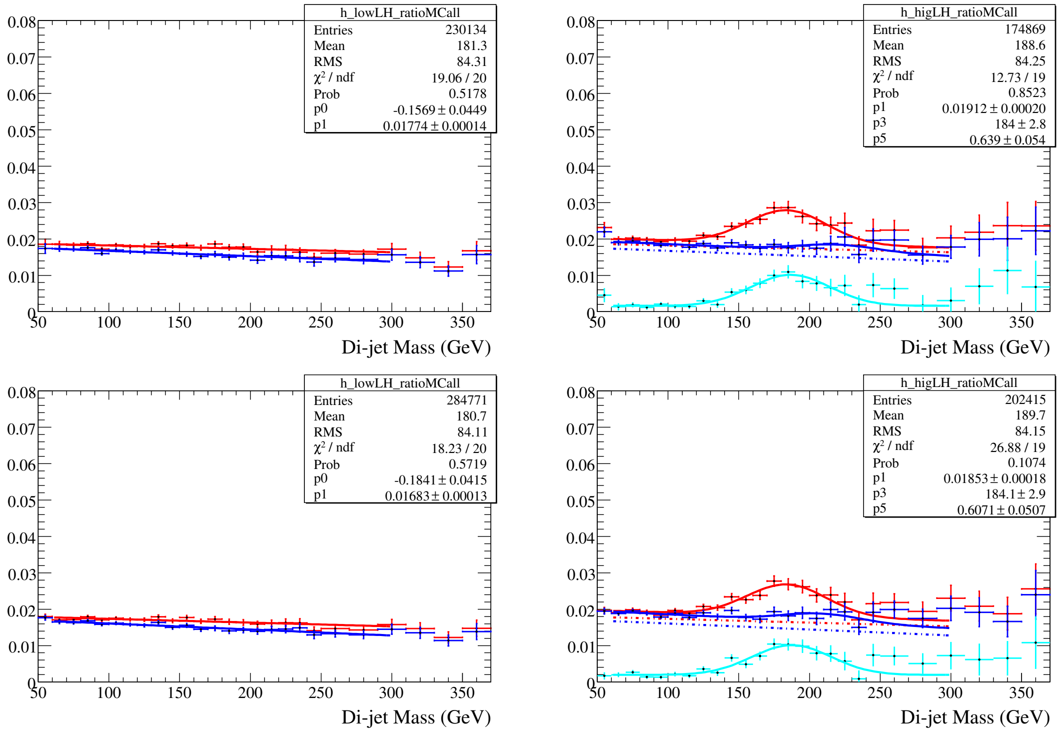

Figure 5.18: The 3/2-tag ratios in the low-discriminant (left) and high-discriminant (right) regions using the optimised NN (top) and likelihood discriminant (bottom) for MC background-only and background-plus-signal hypotheses.

5.8 Systematic Errors

The main sources of systematic uncertainty are listed here, for full details see [22].

5.8.1 Systematics on the Signal

5.8.1.1 From the CTEQ PDFs

Systematic errors due to the choice of CTEQ PDFs were estimated by repeating the signal simulation using each of the 40 eigenvector sets available from the CTEQ collaboration. The variations in the production cross-section for a 100 GeV Higgs were summed quadratically to give a systematic error of ±11.7%.

5.8.1.2 From the Choice of Parameters in MCFM

The choice of renormalisation and factorisation scales used in the MCFM simulation have an impact on the Higgs production cross-section. Both quantities were varied between 0.5 and 2 times the preferred value (the central value was taken from [76, 77]) and the resulting differences in cross-section are taken as the systematic. The systematic errors range from 1.3% for a 90 GeV Higgs to 11.5% at 220 GeV and are combined in quadrature with the those from the CTEQ PDFs described above.

5.8.2 Systematics on the Background

Since the expected background shape is produced from the 2-tag data and the 3/2-tag ratio from MC many of the systematics affecting the 3-tag and 2-tag MC distributions equally cancel out in the ratio and do not affect the final result. The systematics that remain are those that affect samples with different *b*-content differently. They are uncertainties on the background composition, *b*-tagging TRFs and the Level-3 IP-tagger component of the trigger.

5.8.2.1 From the Background Composition, bbj vs. bbb (Systematic 1)

To account for the uncertainty in the background composition derived from the study described previously, the size of the bbj component was varied by 25% for each light jet available to fake the required third *b*-jet (i.e. 25% in the 3-jet channel, 50% in the 4-jet channel and 75% in the 5-jet channel).

5.8.2.2 From Uncertainties in the bbj Sample (Systematic 2)

The fake rate is derived from the negative tag rate as described in the preceding Chapter. This conversion can be approximated by a constant close to unity (~ 0.8). Taking the 2-tag data sample and requiring a third jet to be negatively tagged (and including the correction factor) produces an alternative estimate for the bbj component in the 3-tag distribution. This data sample can then be substituted for the 3-tag bbj MC sample in the constructed background. The background distributions using bbj MC and this data substitute are found to agree within 2% for the 3-jet and 10% for the 4 and 5-jet channels.

5.8.2.3 From the Level-3 IP-tagger b-efficiency (Systematic 3)

A systematic from the Level-3 IP-tagger will only arise if the kinematics of the 2 and 3-tag samples are affected differently by the IP-term in the trigger. To study this an unbiased data sample was collected using a similar trigger to the one used for the analysis. The alternative trigger matches at Level-1 and Level-2 and has a multi-jet requirement at Level-3, but no IP-tagger term. The analysis cuts are applied to the sample and mass pairs are formed as per the final selection. Finally the events from the sample with two and three NN Tight b-tags are passed through the IP-tagger trigger simulation and the ratio of pass rates is formed, parametrised in terms of the di-jet mass. The ratio is then fit with a straight line. The ratio of trigger turn-ons for 2 and 3 offline b-tags is seen to be consistent, within statistical fluctuations, with a constant. The slope from the fit, enhanced by its statistical uncertainty, is then used to calculate the potential deformation of the background shape due to the effect of the IP-tagger trigger term. Since the Level-3 IP-tagger is an event tag, one suitable b-jet is sufficient to fire it. This tagger construction is chosen to ensure high efficiency. Because of this design the difference in performance between two samples which both have at least one high p_T b-jet (in this case a 'mostly-bbj' sample and a 'mostly-bbb' sample) is expected to be small, as is observed.

5.8.2.4 From the Background MC Component Kinematics (Systematic 4)

To investigate the effect of differences in modelling the most significant background component samples, i.e. bbj and bbb, the 3/2-tag MC ratio is recalculated with the 2-tag MC distribution, which is largely bbj, replaced with the 2-tag data. This 3-tag (MC) over 2-tag (data) ratio gives an alternative shape correction. The data 2-tag is used to effectively replace the bbj generator. However the bbj generator is not replaced in the 3-tag distribution and since some of the differences in modelling would cancel, the effect is overestimated. Thus half the effect on the ratio is propagated as a systematic on the final 3-tag shape.

5.8.2.5 From the b-tagging b-efficiency (Systematic 5)

When propagating the systematic errors on the NN b-tagger TRFs two approaches are considered. Firstly the TRFs were shifted up and down uniformly by $\pm 1\sigma$. Secondly the TRFs are skewed by altering the shape by $\pm 1\sigma \rightarrow \mp 1\sigma$. The second procedure produces the larger effect and is used to provide the estimate of the systematic error for the limit setting.

5.8.2.6 From the b-jet Energy Scale (Systematic 6)

The b-jet energy resolution is expected to be less well understood than the light jet resolution. To investigate the effect of this an additional 7% smearing ($\sim \frac{1}{2}$ the experimental resolution) is added to MC b-jets. The effect of this extra smearing is propagated through the variation in the 3/2-tag ratio to a systematic on the final 3-tag shape.

5.8.2.7 Effect of Background Shape Systematics

Section 5.9 details the technique used to set limits in the $m_A - \tan\beta$ parameter space. The limit setting was first carried out with no systematic errors and then with each systematic progressively added in. The resulting effect of the various systematics detailed above can be seen in Figure 5.19. The error from the background composition (systematic 1) is clearly dominant.

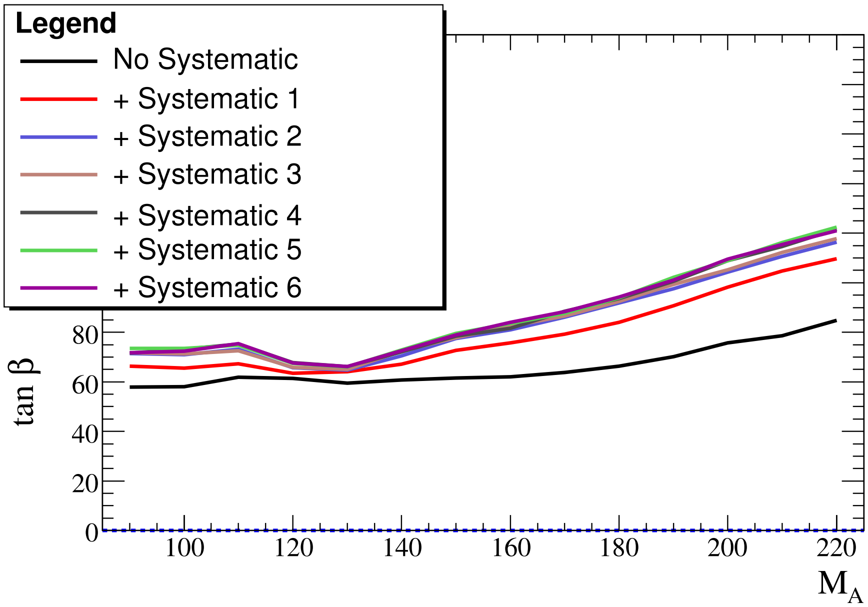

Figure 5.19: The effect on the final $\tan\beta$ limit, calculated at tree level, of progressively adding the systematic errors on the background.

5.9 Final Limits

5.9.1 The Limit Setting Procedure

Limits on $\sigma \times$ BR are intially set at 95% confidence level (CL). These are then converted into limits in $\tan\beta - m_A$ plane, either at tree level (the $2 \times \tan\beta \times 0.9$ enhancement) or in the standard SUSY scenarios quoted in Section 1.3 using `FeynHiggs` [78]. A version of TLimit [79] which had been modified to take into account the unknown level of background (i.e. to use only the background shape) was used to set the limits. The modified version of TLimit is described fully in [80]. TLimit is itself an implementation of MCLimit [81]. Each histogram bin in the 3-tag mass distribution is considered as an independent 'search channel' and then combined to give the final limit for each trial Higgs mass. The modified frequentist confidence level (CL_s) is used where:

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{5.14}$$

where CL_{s+b} is the confidence level for excluding the signal-plus-background hypothesis and CL_b is the confidence level that a background-only fluctuation would give equal to or less than the number of candidates observed. The quantity $1 - CL_b$ is a measure of the compatibility of the data with the background-only hypothesis.

The number of signal and background pairings are varied around their central values using Gaussians with widths equal to the relevant systematic errors. The resulting signal and background levels from a large number of such variants are used to statistically determine the required confidence levels for the signal-plus-background and background-only hypotheses. The signal cross-section is increased until $CL_s \leq 5\%$, giving the 95% exclusion limit for signal.

5.9.2 Low-discriminant Region

As can be seen from the MC 3/2-ratios in Figure 5.18 the effect of a signal in the low-discriminant region is very small. This region can therefore be used as a control region to check that the technique used to estimate the 3-tag background shape is working correctly. The 3/2-tag ratios for data and background-only MC are given in Figure 5.20 and show good agreement. Confidence levels are calculated for the background-only hypothesis and can be seen in Figure 5.21. No excess is observed and so data from the low-discriminant region are said to be consistent with the background-only hypothesis, as expected. The small deviations from the 50% expected value for $1-CL_b$ in the 5-jet channel are due to statistical fluctuations since there are no correlations between neighbouring mass points and the signal statistics for this channel are very low. A cross-check using an injected signal in the 5-jet channel produced a very different $1 - CL_b$ distribution. Since the low-discriminant region behaves as expected the analysis proceeds to the all- and high-discriminant regions.

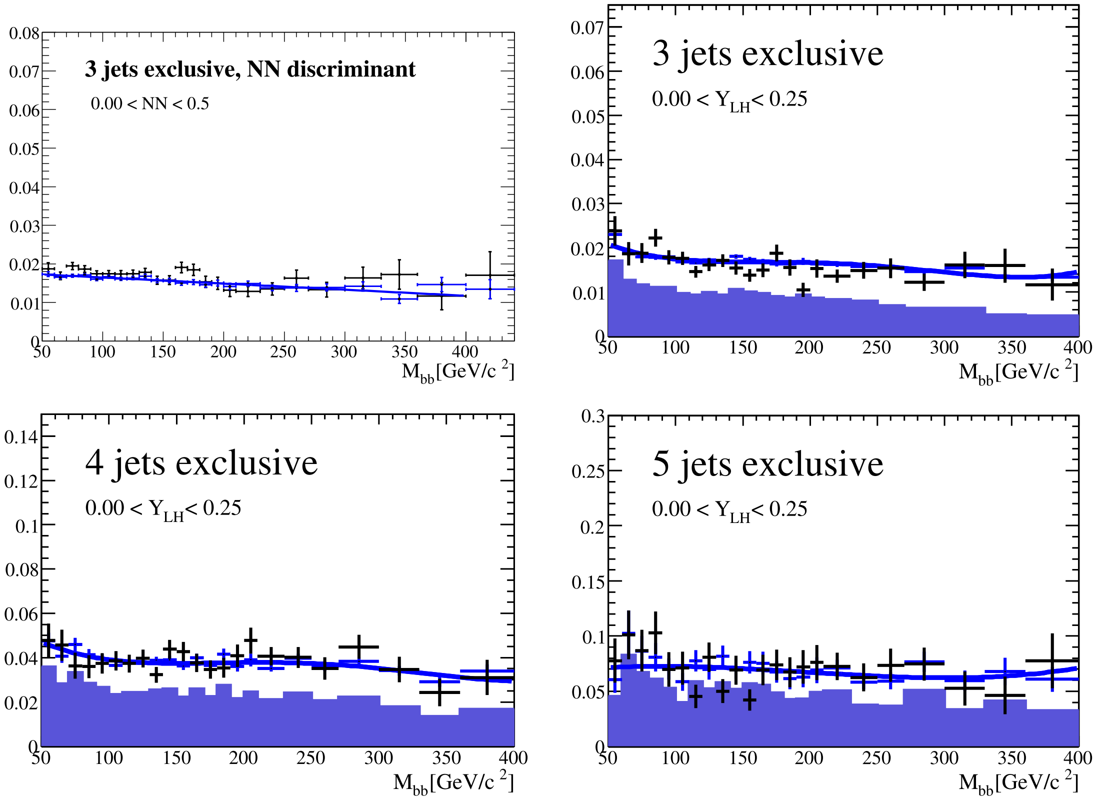

Figure 5.20: The 3/2-tag ratios for the low-discriminant regions, for the NN branch 3-jet channel and likelihood with 3,4 and 5-jets. The blue points are calculatedin each mass bin using MC and the black points are from data. The blue and black lines are polynomial fits to the points in MC and data. The filled purple histogram shows the size of the bbb component in MC (for the likelihood case only).

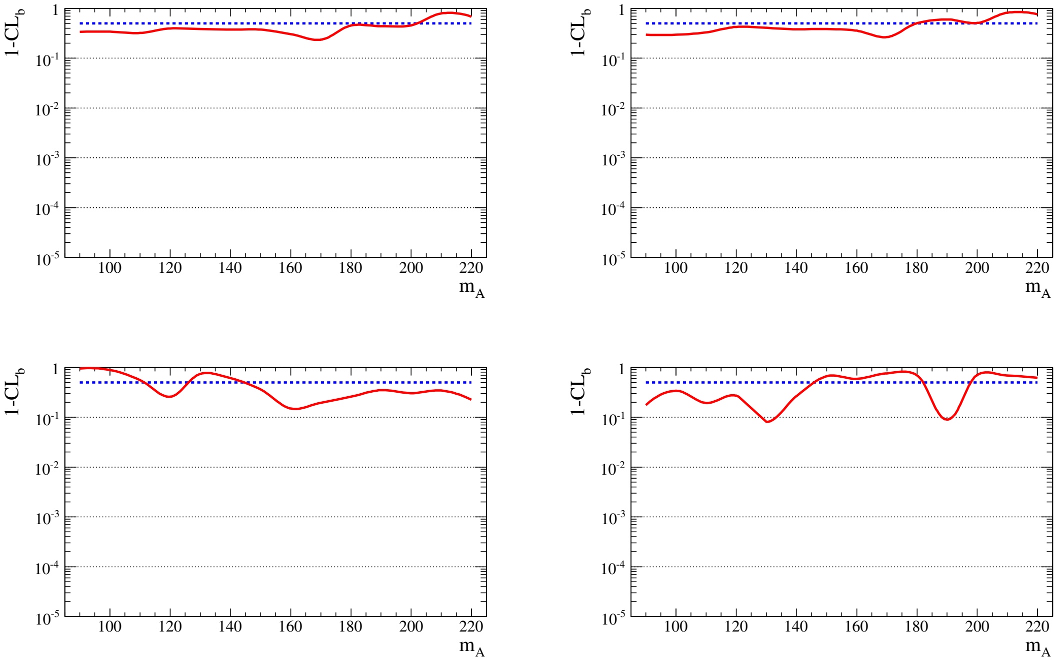

Figure 5.21: The confidence level for the background-only hypothesis in the low-likelihood region for the 3-jet inclusive (top left), 3, 4 and 5-jet exclusive channels (top right, bottom left and bottom right respectively).

5.9.3 No Discriminant Applied

Figure 5.22 shows the limits set and confidence level for the 3-jet inclusive analysis with no multivariate discriminant applied. No significant excess of events is observed and the observed limit agrees well with the expectation.

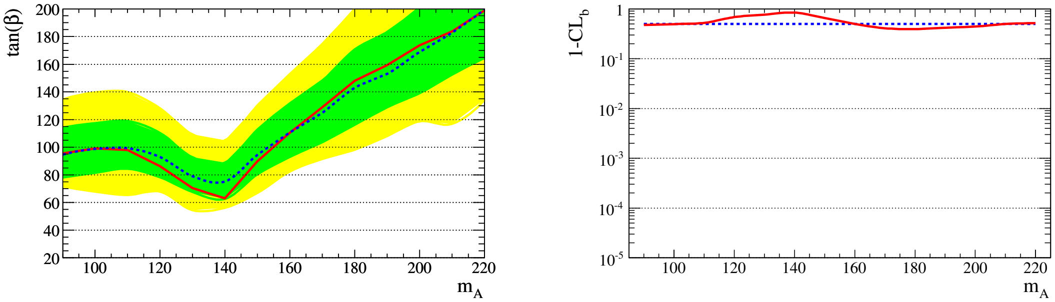

Figure 5.22: Limits on $\tan\beta$ set without using a multivariate discriminant and assuming tree level enhancement (left), observed limits are in red and expected limits dotted black with ±1 and 2σ variations on the expectation given as coloured bands. The confidence level in the background-only hypothesis for the 3-jet inclusive channel is also shown (right).

5.9.4 High-discriminant Region

The 3/2-tag ratios for the high-discriminant region are shown in Figure 5.23. For the region under 150GeV agreement between data and background MC is good, a slight excess in data may be seen above this point. The confidence levels for the background-only hypothesis calculated using the high-likelihood regions of the 3, 4 and 5-jet exclusive and 3-jet inclusive channels are shown in Figure 5.24. As expected from the slight excess observed in the 3/2 tag ratio, the confidence limit at high masses deviates from the 50% expectation.

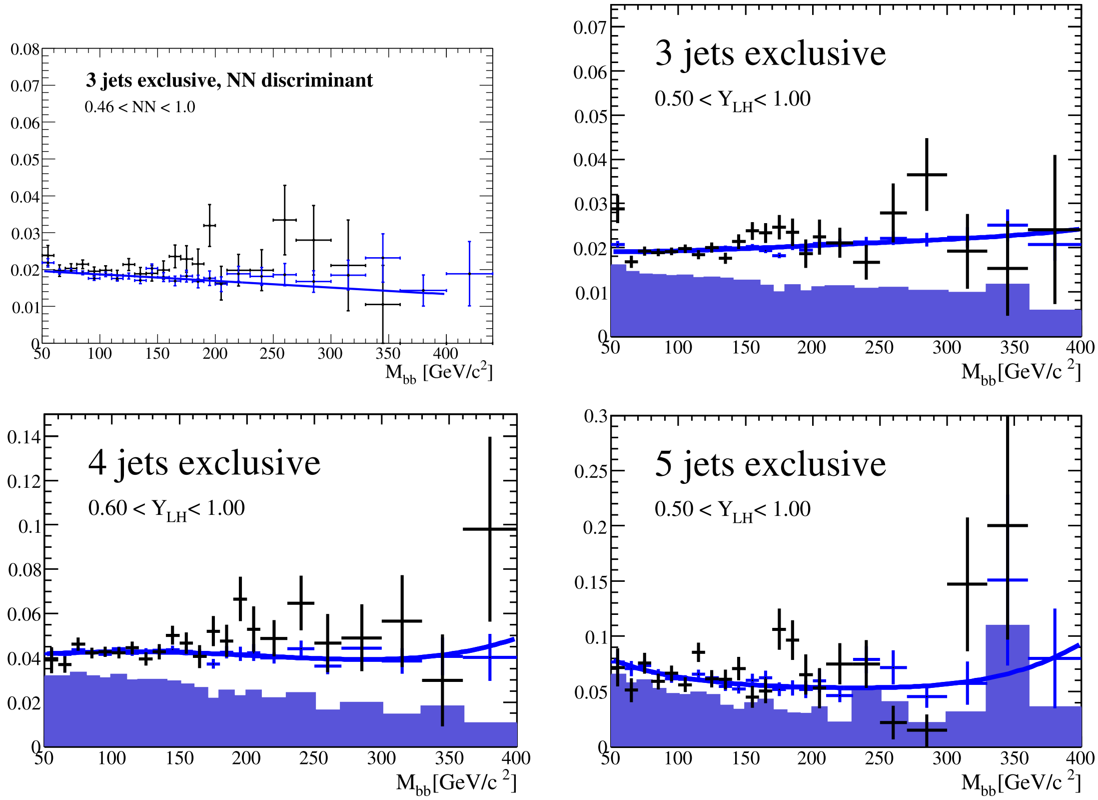

Figure 5.23: The 3/2-tag ratios for the high-discriminant regions, for the NN branch 3-jet channel and likelihood with 3,4 and 5-jets. The blue points are calculatedin each mass bin using MC and the black points are from data. The blue and black lines are polynomial fits to the points in MC and data. The filled purple histogram shows the size of the bbb component in MC (for the likelihood case only).

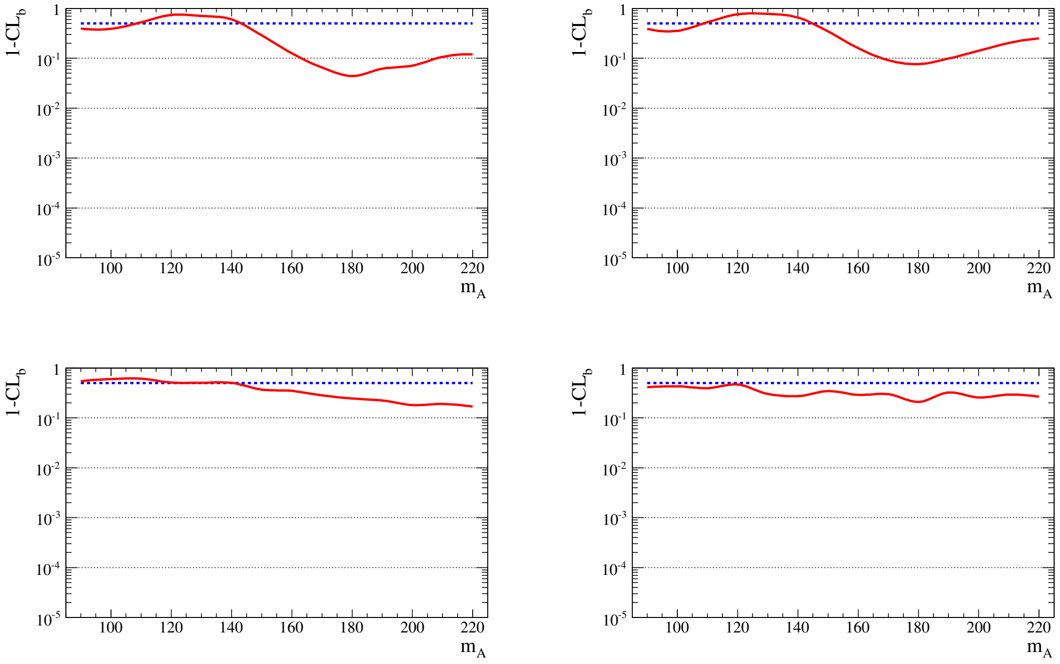

Figure 5.24: The confidence level for the background-only hypothesis in the high-likelihood region for the 3-jet inclusive (top left), 3, 4 and 5-jet exclusive channels (top right, bottom left and bottom right respectively).

5.9.4.1 Limits on the Signal Cross-section and Tree Level $\tan\beta$

Table 5.9 gives the expected and observed limits on the cross-section and$\tan\beta$ as well as the confidence level for the background-only hypothesis and the significance of any deviation from this hypothesis. The 95% confidence limits for the cross-section and $\tan\beta$ are also shown in Figure 5.25.

Again a small excess of events is seen for the higher masses and as a result the observed limit is not as strict as expected. The excess of events is consistent with $m_A \sim 180$ GeV and so the value of $1 - CL_b$ is low for this mass point. The statistical significance of this excess should be reduced by a "look elsewhere" factor since looking for an excess in multiple bins gives an enhanced probability of seeing a significant fluctuation in at least one bin. This factor can be approximated by considering the number of statistically independent mass bins given by the range over twice the typical mass resolution, $(220-90)/2\times20 \approx 3$ [82]. Scaling up $1-CL_b$ by this number allows the excess significance to be recalculated taking into account the probability of finding a fluctuation in at least one independent bin. The rescaled significance for the most significant excess is 1.7σ, as given in Table 5.9.

5.9.4.2 Limits on $\tan\beta$ in the Standard SUSY Scenarios

Figure 5.26 shows the limits on $\tan\beta$ set in the standard SUSY scenarios. The value of μ has a relatively large effect on the limit compared to the other parameters and so two limits for two possible values of μ are shown for the no mixing scenario, with $\mu = \pm200$. The m_h max scenario with $\mu > 0$ is the most unfavourable possibility and does not produce limits less than $\tan\beta = 200$. Thus for the m_h max scenario only the limit for $\mu < 0$ is shown. Figure 5.27 compares these limits with the limits from LEP for the equivalent scenarios.

5.10 Conclusions

This Chapter has presented two generations of analysis in the SUSY Higgs *hb* channel. The preliminary result was the first time the high efficiency NN *b*-tagging described in Chapter 4 had been used in an analysis. The result, presented at ICHEP 06, gave the world's best limit on this process.

Difficulties with the background model following the preliminary result, due in part to changes in the background composition resulting from the NN *b*-tagger's improved performance, led to the introduction of a new analysis method. The final result has been approved by the DØ collaboration and submitted for publication[27].

Higgs boson Mass	$\sigma \times BR$ (pb) Exp.	$\sigma \times BR$ (pb) Obs.	$\tan\beta$ Exp.	$\tan\beta$ Obs.	$1 - CL_b$	signif.
90 GeV/c^2	170^{+72}_{-52}	184	76^{+14}_{-12}	79	39 %	0.3σ
100 GeV/c^2	117^{+48}_{-35}	128	78^{+14}_{-12}	81	38 %	0.3σ
110 GeV/c^2	71^{+29}_{-20}	69	74^{+13}_{-11}	73	52 %	-0.1σ
120 GeV/c^2	41^{+18}_{-9}	34	68^{+13}_{-8}	62	73 %	-0.6σ
130 GeV/c^2	28^{+12}_{-7}	24	66^{+13}_{-9}	62	70 %	-0.5σ
140 GeV/c^2	25^{+11}_{-6}	22	74^{+14}_{-10}	71	60 %	-0.3σ
150 GeV/c^2	21^{+9}_{-6}	26	80^{+16}_{-12}	89	28 %	0.6σ
160 GeV/c^2	17^{+8}_{-4}	26	84^{+18}_{-13}	104	12 %	1.1σ
170 GeV/c^2	14^{+6}_{-3}	24	90^{+19}_{-13}	118	6.4 %	1.5σ
180 GeV/c^2	13^{+5}_{-4}	23	99^{+19}_{-17}	133	4.4 %	1.7σ
190 GeV/c^2	11^{+4}_{-3}	19	105^{+20}_{-19}	138	6.1 %	1.5σ
200 GeV/c^2	9^{+4}_{-3}	17	112^{+24}_{-21}	149	7 %	1.5σ
210 GeV/c^2	8^{+3}_{-2}	13	120^{+25}_{-20}	152	10 %	1.2σ
220 GeV/c^2	7^{+3}_{-2}	12	127^{+27}_{-21}	162	12 %	1.2σ

Table 5.9: Table summarising results, giving the expected and observed limits on $\tan\beta$ assuming tree level enhancement, the expected and observed limits on cross-section times branching ratio, the confidence level for the background-only hypothesis and the significance of any observed excess.

The result is comparable, if not slightly better than, the CDF result in the same channel, despite the CDF result using twice as much integrated luminosity.

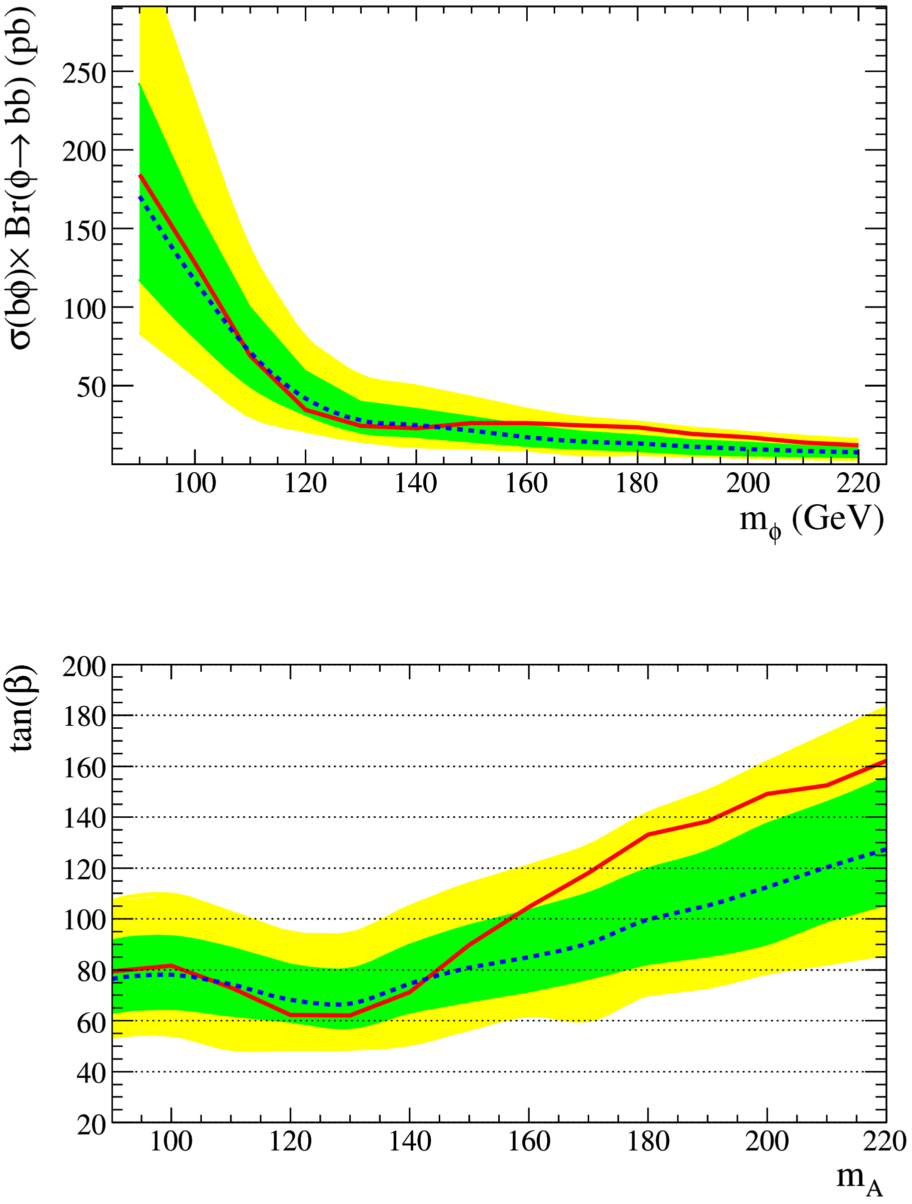

Figure 5.25: The 95% confidence level exclusion limit for: the $(m_A, \tan\beta)$ plane assuming tree level cross-section enhancement (top) and the cross-section for the $hb(b) \rightarrow bbb(b)$ process (bottom). The dotted lines are the expected limits given the background-only hypothesis with ±1 and 2σ bands around the expectation. The observed limits are given by the red curves.

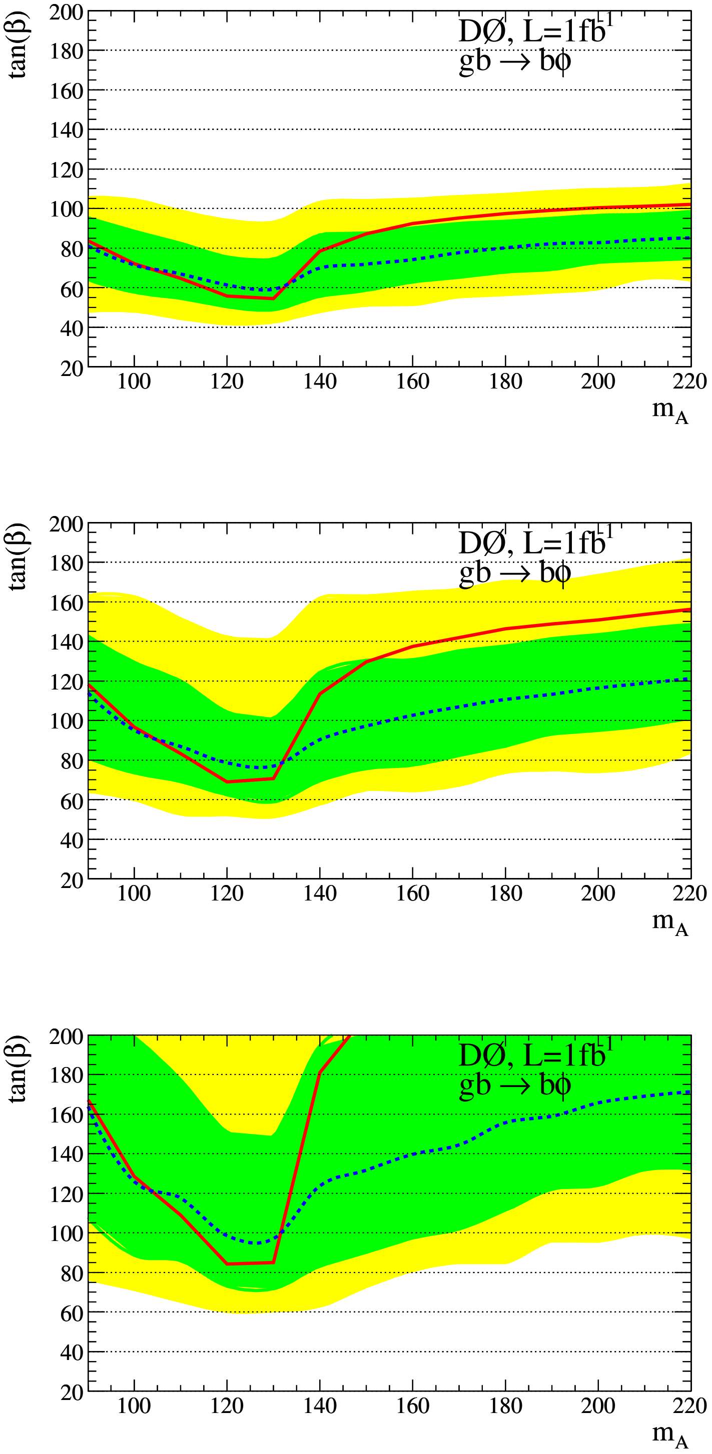

Figure 5.26: The 95% confidence level exclusion limit for: the m_h max scenario with $\mu < 0$ (top), the no mixing scenario with $\mu < 0$ (middle) and the no mixing scenario with $\mu > 0$ (bottom). The dotted lines are the expected limits given the background-only hypothesis with ± 1 and 2σ bands around the expectation. The observed limits are given by the red curves.

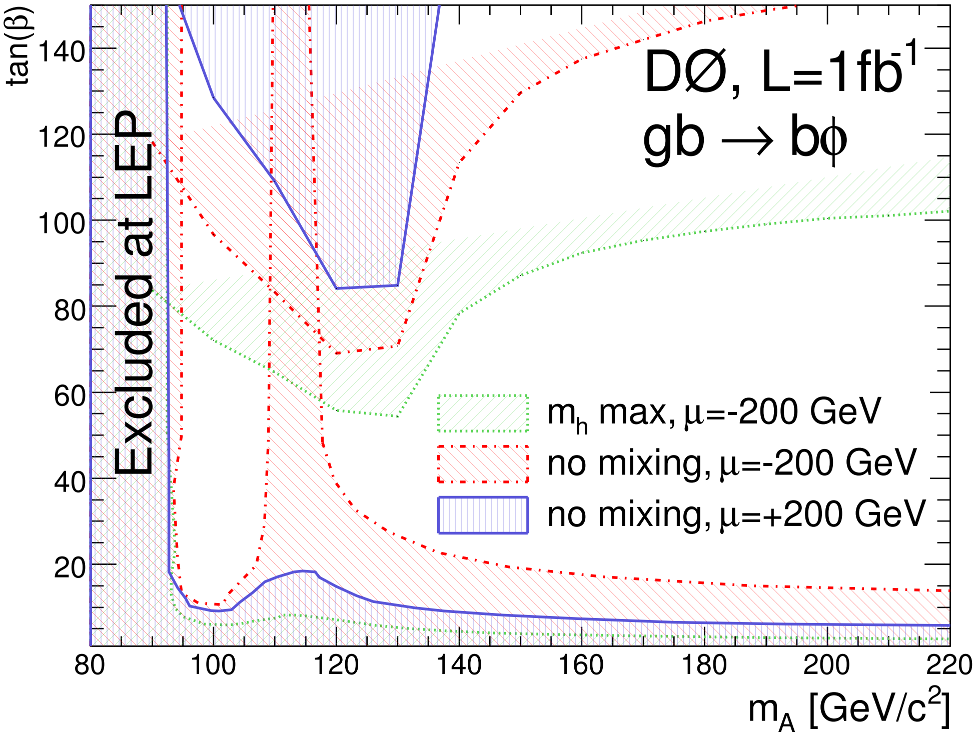

Figure 5.27: The 95% CL exclusion limits for the SUSY scenarios from the present analysis and the equivalent limits from LEP [23]. Below ∼ 92 GeV, and in some scenarios in a limited region around 115 GeV, all values of $\tan\beta$ are excluded by LEP. At other values of m_A LEP can only exclude low values of $\tan\beta$. The results presented here exclude high values of $\tan\beta$.

Chapter 6

Conclusions and Outlook

6.1 Conclusions

Both the DØ detector and the Tevatron are performing well. Data taking efficiency at DØ is nearly 90% close to the maximum workable efficiency for a detector at a hadron collider. The Tevatron has achieved record breaking instantaneous luminosities and is performing in line with its revised design specifications. By the end of Run II DØ is on course to have collected $\sim 8fb^{-1}$ of data.

A proven explanation of the origin of mass is the most significant remaining unknown in the Standard Model. The Higgs mechanism is the most popular theory to provide this explanation. The discovery of the Higgs boson, and thus the confirmation of the existence of the Higgs mechanism within the Standard Model, is one of the most important current goals of the global particle physics community.

Several problems with a universe governed only by the Standard Model may be solved by the introduction of new physics. One effective extension to the Standard Model is Supersymmetry. Supersymmetry allows the 'hierarchy' and 'fine tuning' problems to be solved, and the force couplings to converge at a single grand unification scale. The simplest model of Supersymmetry introduces an additional Higgs doublet resulting in five physical Higgs bosons. Three of these bosons are neutral and have an enhanced coupling to the down-type fermions. In this Minimal Supersymmetric Model the production of a Higgs in association with a b-quark can become the dominant production mode and a decay to two b-quarks the dominant decay mode for all masses. Searches for an excess of events with three or more b-jets can therefore constrain the parameter space of this type of Supersymmetric model, if not make an actual discovery. The efficient identification of b-quarks is essential to distinguish this signal from the large light-jet QCD background, both in the trigger and offline. High quality b-tagging is also important for a large number of other analyses, in particular the low mass SM Higgs search and top analyses.

Given the high rate of unremarkable inelastic collisions triggering is essential at a hadron collider. The Level-3 trigger is the highest level of trigger performing the final, strictest selection of events to be recorded. Level-3 is software based forming a reconstruction of physics objects; tracking and b-tagging are particularly powerful. As the instantaneous luminosity increases Level-3 tools must be continually improved to match the changing detector conditions and retain efficiency/rejection. In parallel new complementary tools are developed to help this process.

Essential maintenance of the Level-3 impact parameter (IP) based b-tagging tool was carried out. The resolution function was corrected to account for improved tracking and increased instantaneous luminosity. As instantaneous luminosities increased physics tools which provide inputs to the IP b-tagger were modified to remain consistent with the constraints of the trigger system. The effect of these changes on the IP b-tagger performance and the effect of increased instantaneous luminosity was investigated. The infrastructure was developed to allow the production of a large b-enhanced data sample for use in trigger studies. The new resolution function and changes to the input tools allowed the efficiency of the IP-tagger to be maintained within the Level-3 time constraint over an order of magnitude increase in instantaneous luminosity.

An extension to the existing IP b-tagging tool to utilise stereo track information was developed. The stereo track errors were re-parametrised, the stereo IP resolution function was measured and the track preselection for the tool was optimised. The power of this tool is fundamentally limited by the resolution of the stereo tracking, but the study none-the-less showed for the first time that stereo information can be used to find b-quarks. It may be possible to make use of the z-IP-tagger output in a future multivariate Level-3 b-tagger.

A secondary vertex based b-tagging tool was co-developed and commissioned for the Level-3 trigger. The parameters used in the algorithm, to select tracks and vertices, were optimised to maximise peak efficiency. A set of operating points was defined for the tool and offered for use in the trigger for the first time. The optimised tool achieved a $\sim 90\%$ efficiency on an offline double-tagged data sample with a $\sim 30\%$ fake rate. This is somewhat better than the IP-tagger which has a fake rate of $\sim 45\%$ for the same efficiency.

Offline b-tagging can again offer powerful background rejection. The offline NN b-tagger was retrained to account for changes to the tracking system and component algorithms. The NN performance was measured on data and cross-checked with closure tests. Appropriate data/MC correction factors were derived. This procedure, known as certification, was reviewed and approved by the DØ collaboration.

For this certification twice as many NN operating points were made available as previously. As well as offering more choice to analysers, this extension of the certification will help suggest if a move to a 'continuous operating point' is of value. The parameter α, in the system eight equations, was measured for the first time. Replacing the previous assigned value of α with a measured value allowed a reduction in the errors on the tag rates. The certification technique was improved to remove the effect of trigger biases in the data samples. An additional error was included in the fake rate to account for different analysis groups' definitions of 'good' jets. This work produced the first version of the NN which was widely adopted throughout DØ. Compared to the individual b-tagging tools previously in use across the Collaboration the NN offers an increase in efficiency of $\sim 30\%$ for fixed fake rate, or a reduction in fake rate to a third at fixed efficiency. The high performance of the NN was instrumental in the single top discovery. The NN-tagger version developed and certified in this thesis has been, and will be, used for all Run IIa DØ results.

Two iterations of the MSSM Higgs search in the $b\phi \rightarrow bb\bar{b}$ channel have been described, each using $\sim 1.0 fb^{-1}$ of data. A critical component of this analysis is the b-tagging at both the trigger level and offline. The earlier version was presented as a preliminary result at ICHEP 2006, giving the then world's best limits on $\tan\beta$. This was also the first use of NN b-tagging in an analysis at a hadron collider. The final result, which has been submitted for publication [27], benefits from an improved description of the background composition, improved signal simulation, a multivariate discriminant, multiple candidate jet pairs and a more robust limit setting technique.

6.2 Outlook

The results from the IP, z-IP and SV based Level-3 b-tagging tools can be combined either directly requiring events to pass multiple cuts or by defining a pass volume in the 3-dimensional space described by the tool outputs. The maximum performance would be achieved by training a NN or using an equivalent multivariate technique.

Layer zero, the additional layer of silicon added to the SMT after Run IIa, has been installed and integrated into the Level-3 tracking algorithm. A new resolution function has been derived using the tools developed in this thesis to allow the IP-based tagger to best exploit this upgrade. With the increased instantaneous luminosity of Run IIb the b-tagging trigger tools are even more important in keeping the Higgs, single top and $t\bar{t}$ trigger rates within the constraints of the data recording system.

The certification of the NN tagger for Run IIb is in progress using the existing certification framework. Parametrising the TRFs with respect to the number of PVs, which is correlated with the instantaneous luminosity, has been suggested as a potential improvement. Expansion of the TRFs in additional parameters should improve the closure tests and therefore reduce the TRF errors. Due to the problem of data/MC differences in the hadron environment extending the NN tagger certification to allow a continuous choice of operating point is non-trivial, however such a development could offer improved background rejection. With a b-tagger certified at a continuous range of operating points combined event tags can be used and should prove more powerful than individual jet tags. Since the performance of the constituent tools vary with jet p_T the possibility of including the jet p_T in the NN is also under investigation. A second NN-tagger to separate b- and c-jets is currently under development, using the same inputs as the existing NN. Another NN b-tagger is under development for muonic jets to combine the information from the lifetime and semi leptonic b-taggers.

The current analysis method has been approved by the Collaboration and will be rapidly applied to a dataset of $\sim 3fb^{-1}$, this result is expected by the end of Summer 2008. Continued improvements to the background model and consequently reduced systematic errors, and the replacement of the likelihood with the NN developed here will further improve the sensitivity. Including events collected with alternative triggers should improve the signal efficiency. The strictest limits on the MSSM parameter space will be set using the combination of the analysis described here with the $b\phi \rightarrow b\tau\overline{\tau}$ and $\phi \rightarrow \tau\overline{\tau}$ channels which have comparable sensitivities to the $b\phi \rightarrow bb\overline{b}$ channel. Ultimately a further factor of two will be gained by the combination of the DØ and CDF results to give the overall Tevatron limits. By the end of Run IIb both experiments are expected to have recorded $\sim 8fb^{-1}$ of data and we expect to be able to exclude $\tan\beta$ values down to ~ 20, if not make a discovery.

References

[1] F. Halzen, A. Martin, *Quarks & Leptons*, John Wiley & Sons, 1984, ISBN 0-471-81187-4.

[2] D. Griffiths, *Introduction to Elementary Particles*, John Wiley & Sons 1987, ISBN 0-471-60386-4.

[3] D. I. Kazakov, *Beyond the Standard Model (in Search of Supersymmetry)*, hep-ph/0012288 v2.

[4] W. Yao et al. (Particle Data Group), *Supersymmetry*, J. Phys. G **33**. 1 (2006).

[5] S. Dawson, *SUSY and Such*, hep-ph/9612229.

[6] M. Grunewald, *Precision Tests of the Standard Model*, hep-ex/0511018v1.

[7] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)

[8] S. Glashow, J. Iliopoulos, L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2**, 1285-1292 (1970).

[9] P. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Phys. Lett. **12**, 132 (1964).

[10] P. Higgs, *Broken Symmetries and the Masses of the Gauge Bosons*, Phys. Lett. **13**, 508 (1964).

[11] P. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145**, 1156 (1966).

[12] J. Goldstone, A. Salam, S. Weinberg, *Broken Symmetries*, Phys. Rev. **127**, 965 (1962).

[13] K. Riesselmann, *Limitations of a Standard Model Higgs Boson*, hep-ph/9711456.

[14] T. Appelquist, M. Chanowitz, *Unitarity Bound on the Scale of Fermion Mass Generation*, Phys. Rev. Lett. 59, 2405.

[15] The LEP Electroweak Working Group, *A Combination of Preliminary Electorweak Measurements and Constraints on the Standard Model*, `http://lepewwg.web.cern.ch/LEPEWWG/plots/winter2008/`.

[16] The LEP Working Group for Higgs Boson Searches, *Search for the Standard Model Higgs Boson at LEP*, Phys. Rev. Letters B **565**, pp. 61-75 (2003).

[17] M. Carena, S. Heinemeyer, C. Wagner, G. Weiglein, *Suggestions for Benchmark Scenarios for MSSM Higgs Boson Searches at Hadron Colliders*, hep-ph/0202167.

[18] M. Spira, *Higgs boson production and decay at the Tevatron*, hep-ph/9810289.

[19] A. Haas, A. Kharchilava, M Michaut, J. Rani, T. Scanlon, B. Tuchming *DØ Search for Neutral Higgs Bosons at High* $\tan\beta$ *in Mulit-jet Events Using p14 Data*, DØ Note 4671.

[20] B. Ananthanarayan, G. Lazarides, Q. Shafi, *Top Mass Predicions from Supersymmetric Grand Unified Theories*, Phys. Rev. D **44**, 1613-1615 (1991).

[21] V. Barger, C. Kao, *Implications of New CMB Data for Neutralino Dark Matter*, hep-ph/0106189.

[22] F. Couderc, B. Tuchming, G. Davies, J. Hayes, P. Jonsson, S. Robinson, T. Scanlon, DØ Search for Neutral Higgs Bosons at High $\tan\beta$ in Mulit-jet Events Using p17 Data, DØ Note 5341.

[23] ALEPH, DELPHI, L3, OPAL Collaborations. The LEP Working Group for Higgs Boson Searches, *Search for Neutral MSSM Higgs Bosons at LEP*, Eur.Phys.J. C47 (2006) 547-587, hep-ex/0602042v1.

[24] F. Couderc, B. Tuchming, M. Michaut, G. Davies, P. Jonsson, S. Robinson, T. Scanlon, *DØ Search for Neutral Higgs Bosons at High* $\tan\beta$ *in Mulit-jet Events Using p17 Data*, DØ Note 5202.

[25] The DØ Collaboration, *Search for Neutral Higgs Boson Production in the Decay* $h \rightarrow \tau(\mu)\tau$ *with the DØ Detector at* $\sqrt{s} = 1.96\ TeV$, DØ Note 5331.

[26] The CDF collaboration, CDF Note 8676.

[27] The DØ collaboration, *Search for neutral Higgs bosons in multi-b-jet events in ppbar collisions at sqrt(s)=1.96 TeV*, Phys. Rev. Lett., arXiv:0805.3556; Fermilab-Pub-08/142-E.

[28] Fermi National Accelerator Laboratory (Fermilab), `http://www.fnal.gov`

[29] Fermilab Beams Division, RunII Handbook, `http://www-bd.fnal.gov/runII/index.html`

[30] The CDF Collaboration, *Observation of Top Quark Production in Pbar-P Collisions*, hep-ex/9503002.

[31] The DØ Collaboration, *Observation of the Top Quark*, Phys. Rev. Letters **74**, 2632.

[32] V.M. Abazov, et al. (D0 Collaboration), *The Upgraded DØ Detector*, Nucl. Instrum. and Methods A **565**, 463-537 (2006).

[33] R. Lipton, *Layer 0 performance*, DØ workshop, August 2006.

[34] The DØ collaboration, *Beam Tests of the DØ Uranium Liquid Argon End Calorimeter*, Nucl. Instr. and Methods A 324 53, Fermilab-Pub-92/162-E.

[35] A. Schwartzman, C. Tully, *Primary Vertex Reconstruction by Means of Adaptive Vertex Fitting*, DØNote 4918.

[36] A. Schwartzman, Y. Peters, H. Greenlee, *Adaptive Primary Vertex Certification in p17*, DØNote 5192.

[37] G. Blazey, et. al., *Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop*, hep-ex/0005012.

[38] N. Makovec, J. Grivaz, *Shifting, Smearing and Removing Simulated Jets*, DØ Note 4914.

[39] A. Schwartzman, M. Narain, *Track-Jet Studies Using a 3-Dimensional Simple Cone Jet Algorithm*, DØNote 3919.

[40] C. Gerber, E. Chabalina, G. Garzon, *Taggability in Pass2 p14 Data*, DØNote 4995.

[41] C. Clement et. al., *MuonID Certification for p14*, DØ Note 4350.

[42] D. Bloch, B. Clément *Update of the JLIP b-tagger Performance in p14/pass2 with JES 5.3*, DØ Note 4824.

[43] R. Demina, A. Khanov, F. Rizatdinova, *b-Tagging with Counting Signed Impact Parameter Method*, DØNote 4049.

[44] A. Schwartzman, M. Narain, *b-quark Jet Identification with Secondary Vertex Reconstruction using DØReco Version p14*, DØNote 4414.

[45] K. Hanagaki, J. Kasper, *Identification of b-jet by Soft Muon*, DØNote 4867.

[46] R. Illingworth, *Raw Data Format and Unpacking for the SMT*, DØNote 3829.

[47] D.Bauer, R.Illingworth, *Level 3 Trigger: Clustering in the SMT*, DØNote 3822.

[48] L3 CFT Unpacking Tool, `http://www-d0.fnal.gov/d0dist/dist/packages/l3fcftunpack/p16-br-04/`.

[49] D. Whiteson, *Global Track Finding at Level 3*, DØNote 3808.

[50] Imperial College DØ Group, *A Level 3 CFT Tracking Tool*, DØNote 3779.

[51] C. Barnes, *Development of vertexing and lifetime triggers and a study of Bs mixing using hadronic decays*, PhD Thesis.

[52] A. Khanov, *HTF: Histogramming Method for Finding Tracks. The Alogrithm Description*, DØ Note 3778.

[53] C. Barnes, R. Beuselinck, P. Jonsson, *Vertexing for the Level 3 Trigger*, DØ Note 4271.

[54] M. Narain, F. Stichelbaut, *Vertex Reconstruction using the Impact Parameters Technique*, DØ Notte 3560.

[55] V. Buscher, *Calorimeter Clustering Tool for the DØ Level 3 Trigger at Run II*, `http://www-d0.fnal.gov/computing/algorithms/level3/jets/l3fcalcluster.ps`.

[56] Level 3 Jet Tool Documentation, `http://www-d0.fnal.gov/computing/algorithms/level3/jets/L3TJet_overview.html`.

[57] DØ Trigger Simulator (TrigSim) `http://www-d0.fnal.gov/computing/trigsim/trigsim.html`

[58] T. Sjöstrand, L. Lönnblad, S. Mrenna and P. Skands, *Physics and manual*, hep-ph/0308153. Pythia 6.319 and 6.323 are used.

[59] T. Scanlon, S. Robinson, *A Secondary Vertex b-Tagging Tool for the Level-3 Trigger*, DØNote 5292.

[60] T.Scanlon, *b-Tagging and the Search for Neutral Supersymmetric Higgs Bosons at DØ*, FERMILAB-THESIS-2006-43.

[61] T. Scanlon, *A Neural Network b-tagging Tool*, DØ Note 4889.

[62] M. Anastasoaie, S. Robinson, T. Scanlon, *Performance of the NN b-tagging Tool on p17 Data*, DØ Note 5213.

[63] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307**, 001 (2003) [arXiv:hep-ph/0206293].

[64] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, ISBN 0-13-273350-1

[65] TMultiLayerPerceptron `http://root.cern.ch/root/html/TMultiLayerPerceptron.html`

[66] K. Hornik et al., *Multilayer Feedforward Networks are Universal Approximators*, Neural Networks, Vol. 2, p359-366 (1989).

[67] M. Avriel, *Nonlinear Programming: Analysis and Methods*, Dover Publishing, ISBN 0-486-43227-0.

[68] R.Fletcher, *Practical Methods of Optimization*, second edition, Wiley (1987).

[69] J. Ortega, W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics in Applied Mathematics, SIAM, 2000. ISBN 0-89871-461-3.

[70] B. Clement, D. Bloch, D. Gele, S. Greder, A.C. Le Bihan, I. Ripp-Baudot, *SystemD or how to get signal, backgrounds and their efficiencies with real data*, DØ Note 4159.

[71] TMinuit, `http://root.cern.ch/root/html/TMinuit.html`.

[72] T. Scanlon, M. Anastasoaie, *Performance of the NN b-tagging Tool on Pass2 Data*, DØ Note 4890.

[73] CTEQ Collaboration, `http://www.phys.psu.edu/~cteq/`.

[74] Common Analysis Format, `http://www-d0.fnal.gov/Run2Physics/cs/caf/`

[75] J.M. Campbell, R.K. Ellis, `http://mcfm/`.

[76] S. Dittmaier, K. Kramer and M. Spira, hep-ph/0309204.

[77] S. Dawson, C. Jackson, L. Reina and D. Wackeroth, Phys. Rev. D **69** (2004) 074027.

[78] S. Heinemeyer, W. Hollik and G. Weiglein, Eur. Phys. **J**. C **9**, 343; Comput. Phys. Commun. **124**, 76; G. Degrassi *et al.*, Eur. Phys. **j**. C **28**, 133; M. Frank *et al.*, JHEP **0702**, 047. `FeynHiggs` version 2.6.3 is used.

[79] TLimit, `http://root.cern.ch/root/html/TLimit.html`.

[80] F. Couderc, *Limit setting in the presence of an unknown level of background*, DØ Note 5327.

[81] T. Junk, *Confidence Level Computation for Combining Searches with Small Statistics*, Nucl. Instrum. Meth. A434 (1999) 435-443.

[82] The LEP Collaborations, Phys Lett B 565 (2003) 61.