UNIVERSITY OF CALIFORNIA

Santa Barbara

Measurement of B Meson Semileptonic Decay Form Factors

A Dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Anders Ryd

Committee in charge:
Professor Jeffrey Richman, Chairperson
Professor Rollin Morrison
Professor Robert Sugar

June 1996

The dissertation of Anders Ryd is approved

Committee Chairperson

June 1996

ACKNOWLEDGEMENTS

First I would like to thank Jeff Richman, my advisor, who encouraged me to stay at UCSB and pursue graduate study in experimental particle physics. I'm very glad that I decided to work for Jeff since I feel that I have learned a lot during the years I worked with him.

I would also like to thank the Rollin Morrison and Mike Witherell for their expertise on form factors which has been a great help.

Other people who have contributed greatly to my work and that I would like to thank are Ron Poling, Ritchie Patersson, Duncan Gibaut, and Xu Fu.

The University of Minnesota and the University of Hawaii groups generously helped by supplying computer resources for Monte Carlo generation, which I'm grateful for.

For the work on the Billoir fitter I'd like to thank Rob Kutschke whom I worked closely with.

I would like to thank all my friends that have been around for the time I spent in graduate school both at UCSB and Cornell. In particular, I would like to thank Mingming Wu, Matt Britton, Alain Bellerive, David Lange, Doug Roberts, Chris Korte, Alice Bean, Jeff Gronberg, and Dave Flynt.

A special thanks also to my parents who always have supported me through my educational endeavors.

CURRICULUM VITAE

Born: November 18, 1968, Ask, Sweden.

Education:

Sept. 1991 - June 1996:	Ph.D. Physics University of California, Santa Barbara.
Sept. 1987 - Sept. 1991:	B.S. Physics, University of Lund, Sweden.
June 1989 - June 1990:	Military service in Sweden.

PUBLICATIONS

"Measurement of the Form Factors for $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$", J. Doboscq *et al.*, Phys. Rev. Lett. **76**, 3899 (1996).

"Measurement of the Form Factors for $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$", A. Ryd, Proceedings of the 8th meeting of the Division of Particles and Fields of the American Physical Society, page 847 (1994).

"Measurement of the Form Factors for $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$", A. Ryd, to be published in the proceedings from 2nd Rencontres du Vietnam, Saigon, Vietnam (1995).

Measurement of B Meson Semileptonic Decay Form Factors

by

Anders Ryd

Abstract

This work reports on a measurement of the semileptonic form factors that govern the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$. This measurement is based on the CLEO II data set, containing approximately 2.6×10^6 $B\bar{B}$-pairs. The D^{*+} is reconstructed in the channel $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$ or $D^0 \to K^-\pi^+\pi^0$. This analysis used 940 reconstructed events with an estimated background contamination of 17%. The measurement consists of performing a likelihood fit in the four kinematic variables that describe the decay for the three parameters R_1, R_2, and $\rho^2_{A_1}$. Previous measurements have either measured form factor ratios assuming a q^2 dependence of the form factors or they have measured the q^2 dependence assuming the form factor ratios. The results obtained for the form factor ratios are $R_1 = 1.18 \pm 0.30 \pm 0.12$ and $R_2 = 0.71 \pm 0.22 \pm 0.07$. The form factor slope is found to be $\rho^2_{A_1} = 0.91 \pm 0.16 \pm 0.06$. These results are consistent with calculated corrections to the heavy quark symmetry limit.

Contents

List of Figures

List of Tables

Part I

Measurement of the $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ form factors

Chapter 1

Introduction

The study of semileptonic decays provides a rich source of knowledge for determining fundamental parameters of the standard model. In particular, semileptonic decays have helped, and will continue to help, determine the parameters of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [1]. The study of semileptonic B decays provides a means to measure the CKM matrix elements $|V_{cb}|$ and $|V_{ub}|$. As in the determination of other elements of the CKM matrix, the extraction of the parameters is complicated by the effects of strong interactions of the hadrons that confine the quarks. In the determination of $|V_{ud}| = 0.9744 \pm 0.0010$ using beta decays [2], the effects of strong interactions can be reliably calculated using isospin, a symmetry under the exchange of u and d quarks. This symmetry arises from the small difference between the u and the d quark masses. In particular, corrections to the symmetry limit are of the order $(m_u - m_d)/\Lambda_{\rm QCD}$, which is a few percent.

Without the help of isospin symmetry, the theoretical uncertainty on the extracted values of the CKM matrix elements are much larger. As an example, consider V_{cs}. Here the determination is completely limited by the theoretical uncertainty; experiments measure $|V_{cs}f_+(0)| = 0.70 \pm 0.05$ [2] while $|V_{cs}| = 1.0 \pm 0.20$ is limited by the theoretical prediction of $f_+(0)$.

However, for the process of interest in this dissertation, $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$, in which a b-quark decays semileptonically to $c\ell\nu$, a new symmetry comes to the rescue. This symmetry, heavy quark symmetry (HQS), is a symmetry under the exchange of heavy quarks. In the limit that the b and c quarks can be considered heavy, this allows model independent extraction of the CKM matrix element $|V_{cb}|$ from the decays $B \to D\ell\nu$ and $B \to D^*\ell\nu$ [3]. In the symmetry limit ($m_Q \to \infty$) the spin of the heavy quark in the meson decouples from the dynamics. Thus, the symmetry group is SU($2N_h$), where N_h is the number of heavy quark flavors. This symmetry leads to relations between the form factors describing the decays $B \to D\ell\nu$ and $B \to D^*\ell\nu$. The work presented in this dissertation provides a test of the predictions of HQS and gives confidence in the extraction of $|V_{cb}|$ using relations based on HQS.

1.1 B semileptonic decays

B semileptonic decays are studied primarily at the $\Upsilon(4S)$ and at the Z^0 resonance in e^+e^- collisions. The advantage at the $\Upsilon(4S)$ is the production of two B mesons, and nothing more, almost at rest in the laboratory frame. This gives clean events with a background of nonresonant $q\bar{q}$ ($q = u, d, s, c$) events. It is easy to obtain a clean sample of these background events by collecting data below the energy for producing $\Upsilon(4S)$. The two experiments ARGUS [4] and CLEO [5] have collected data (2×10^5 and 3×10^6 $B\bar{B}$ pairs respectively) at the $\Upsilon(4S)$ and have made enormous progress in the understanding of the physics of B mesons.

The experiments at LEP, in particular ALEPH and OPAL, have pursued successful programs in the area of B-physics, which was not the primary goal of the LEP experiments. The major advantage of using data collected at the Z^0 is that the B mesons appear as separate vertices when they decay.

The experiments CDF and D0 also have programs in B-physics. Their data are collected in $p\bar{p}$ collisions at $\sqrt{s} = 1,800$ GeV. The disadvantage here is that a small fraction of the events contains b-quarks and that the events have very high multiplicity. Triggering on events containing b quarks is hard except in the subset that produces leptons with a high transverse momentum.

One of the most well measured quantities in B-physics is the inclusive semileptonic branching fraction

$$B(B \to X\ell\nu) = (10.49 \pm 0.50)\% \tag{1.1}$$

measured using dilepton events [6]. The dilepton technique uses the charge correlation between the two leptons to separate the primary and secondary lepton spectra. Using the measured B meson lifetime this inclusive measurement can be used to extract $|V_{cb}|$ [1]. There is a largetheoretical uncertainty

[1] To extract $|V_{cb}|$ the inclusive branching fraction needs to be corrected for $b \to u\ell\nu$. This is a small correction of approximately 2 percent.

in predicting the inclusive rate. Using the inclusive model ACCMM [7] $|V_{cb}|$ is found to be

$$|V_{cb}| = (40 \pm 1 \pm 4) \times 10^{-3}. \tag{1.2}$$

In D semileptonic decays, the two exclusive channels $D \to K\ell\nu$ and $D \to K^*\ell\nu$ account for almost all of the semileptonic rate. In semileptonic B decays the situation is different; the two largest exclusive channels are [8, 9]

$$B(B \to D\ell\nu) = (2.0 \pm 0.6)\% \tag{1.3}$$

and

$$B(B \to D^*\ell\nu) = (4.5 \pm 0.5)\%, \tag{1.4}$$

which account for only about $60 - 70\%$ of the semileptonic width. Recent measurements [10, 11] indicate that there are significant rates to excited and nonresonant D states, throughout this dissertation called D^{**}s. However, a complete picture of which exclusive channels make up the semileptonic width has still not emerged.

The largest exclusive semileptonic channel, in fact, the largest exclusive decay channel of B mesons, is $B \to D^*\ell\nu$, with a branching ratio of $\approx$ 5%. This mode provides a suitable laboratory for the detailed study of the dynamics that the measurement of the form factors requires. In the work presented in this dissertation the channel $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$, $D^{*+} \to D^0\pi^+$ is used since it provides the cleanest mode due to the charged slow pion. The decay channels $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^+\pi^0$ are used to reconstruct the D^0. Both electrons and muons are used for the lepton candidates.

1.2 Form factors in semileptonic decays

The semileptonic decay of a meson is mediated by a virtual W [12], a charged vector boson. In this process a quark (in decays of $\bar{B}$ mesons a b-quark) decays to a lighter quark, e.g., a c-quark, and emits a virtual W. The virtual W decays to a lepton-neutrino pair in the semileptonic decay, as shown in Figure 1.1. As discussed in Section 2.2 the form of the weak current, the coupling of the W to the quarks and the leptons, is well known in the standard model. However, theoretical predictions of the matrix elements in semileptonic decays are complicated by the effects of strong interactions. These interactions are not calculable perturbatively; attempts have been made to calculate the effects of the strong interactions using lattice QCD or QCD sum rules [13]. However, a consistent picture has not yet emerged from these approaches.

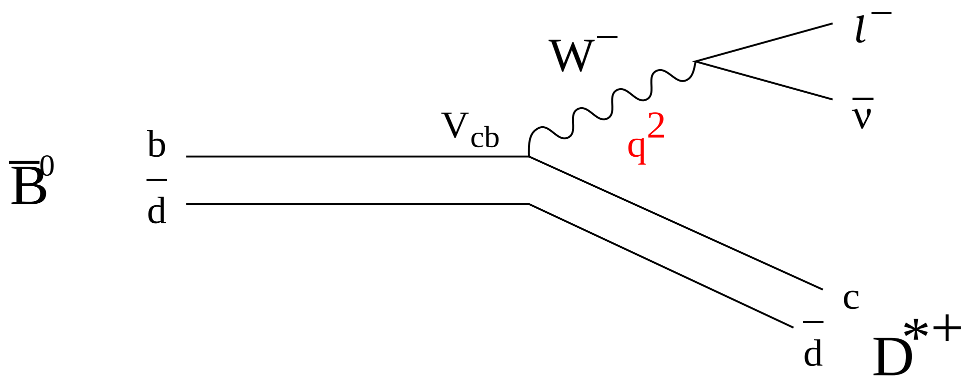

Figure 1.1: Quark level diagram for the semileptonic decay.

The effects of the strong interactions can be parametrized in terms of a set of form factors. The number of form factors needed to describe a particular decay depends on the spins of the initial and final state mesons. In the decay this work is focused on — (pseudo)scalar to vector meson – there are four form factors. However, for negligible lepton masses, $\ell = (e, \mu)$, only three form factors contribute to the rate. These form factors are functions of a single scalar variable, often taken to be q^2, the mass squared of the virtual W, as discussed in Section 2.2.

There are several different ways of defining the form factors. The simplest approach that directly relates the form factors to the weak hadronic current is to use invariant form factors. In the semileptonic decay of a pseudoscalar to a pseudoscalar, e.g., $B \to D\ell\nu$ the weak hadronic current is of the form

$$\langle D(p')|c\gamma^\mu(1-\gamma^5)\bar{b}|B(p)\rangle = f_+(q^2)(p+p')^\mu + f_-(q^2)(p-p')^\mu. \quad (1.5)$$

This current is constructed from the requirement that it transforms as a vector under Lorentz transformations. The current is constructed from the available 4-vectors, p and p'. Expressions such as $(p+p')^\mu$, that transform as 4-vectors under Lorentz transformations can be multiplied by an arbitrary function that only depends on Lorentz scalars. There is only one independent Lorentz scalar in this process, which is taken to be $q^2 = (p-p')^2$. The functions, $f_+(q^2)$ and $f_-(q^2)$ in the example above, are the invariant form factors. Note that the definition is not unique, e.g.,

$$\langle D(p')|c\gamma^\mu(1-\gamma^5)\bar{b}|B(p)\rangle = f(q^2)p^\mu + f'(q^2)p'^\mu \quad (1.6)$$

is also an perfectly valid and general parametrization of the current. There is, however, a unique mapping between the two sets of form factors.

In the decay of a pseudoscalar to a vector meson the current is also required to be linear in ε the polarization of the vector meson. The presence of another vector, ε^{μ}, in the process gives rise to a richer structure of the current, as discussed in Section 2.2.

The invariant form factors are related to the helicity amplitudes in Section 2.2. The helicity amplitude gives the amplitude for a decay with a given helicity of the W. It is convenient to express the decay rate using the helicity amplitudes. The helicity amplitudes are also fairly intuitive, i.e., some properties of the helicity amplitudes can be argued from simple physical considerations, as discussed in Section 2.3.4.

1.3 Heavy quark symmetry

In the limit that the mass of the quarks in a semileptonic transition, e.g., the b and c quarks in $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$, can be considered infinite, new symmetries arise in the Lagrangian of the strong interactions [14, 15]. The flavor symmetry under the exchange of heavy quarks and the spin symmetry give relations between the form factors that parametrize the effects of the strong interactions. This is discussed in Section 2.6.1. It is important to note that heavy quark symmetry, HQS, does not predict the form factors, it relates them to one universal form factor, the Isgur-Wise function, $\xi(q^2)$. However, it does give the absolute normalization of the Isgur-Wise function at the zero-recoil configuration, the configuration in which the final state meson is at rest in the parent meson rest frame.

The b and the c quarks have finite masses and the effects of the finite masses can be taken into account in a systematic way using heavy quark effective theory, HQET. HQET is a systematic expansion of the QCD Lagrangian in the expansion parameter E/m_Q where m_Q is the heavy quark mass and E is a relevant energy scale. For some purposes E can be taken to be Λ_{QCD}, which is approximately a few hundred MeV. These corrections have been calculated and some of the results that are relevant to the measurement of the form factors in $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ are summarized in Section 2.6.2.

1.4 Measurement of the form factors in $B \to D^*\ell\nu$

Based on the predictions from HQS, the form factors governing the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ are parametrized in terms of three parameters: R_1, R_2, and $\rho^2_{A_1}$. There are two form factor ratios, R_1 and R_2, and one parameter, $\rho^2_{A_1}$, describing the q^2 dependence of the form factors. Since this analysis only uses electrons and muons, which are essentially massless, i.e., $m_\ell^2/q^2 \ll 1$ for most of phase space, there are only three form factors contributing to the rate.

The measurement of the form factors consists of performing a likelihood fit for the parameters described above in the space of the four kinematic variables that describe the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ followed by $D^{*+} \to D^0\pi^+$; the kinematics of this decay are described in Section 2.1.

Previous measurements of the form factors were limited to measuring either derived quantities or making assumptions about either the form factor slope or the form factor ratios. The D^* polarization, $\bar{A}_{\rm pol}$, and the lepton forward-backward asymmetry in the W rest frame, $\bar{A}_{\rm FB}$, have been measured in the process $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$. Another measurement related to this form factor study is the measurement, inspired by HQS, of the q^2 dependence of the Isgur-Wise function.

CLEO [16] has measured $\bar{A}_{\rm FB} = 0.14 \pm 0.06 \pm 0.03$ for a lepton energy above 1.0 GeV. ARGUS [17] has also performed this measurement but quotes the result over the full lepton energy range, $\bar{A}_{\rm FB} = 0.20 \pm 0.08 \pm 0.06$. They have also measured the D^* polarization, $\bar{A}_{\rm pol} = 1.1 \pm 0.4 \pm 0.2$. These quantities are derived from the form factors measured in this dissertation to allow comparison with these previous measurements.

Both CLEO [9] and ARGUS [17] have measured the form factor slope $\hat{\rho}^2$ using the assumption that $R_1 = R_2 = 1$. CLEO found $\hat{\rho}^2 = 0.84 \pm 0.12 \pm 0.08$ and ARGUS found $\hat{\rho}^2 = 1.17 \pm 0.24 \pm 0.06$. The form factor slope found in these measurements can not be directly compared to $\rho^2_{A_1}$, but in Section 7.3 a relation between the two form factor slopes is presented that allows a comparison.

An attempt is made in Reference [16] to extract the form factor ratios by combining the information from the measurements of $\bar{A}_{\rm FB}$, $\bar{A}_{\rm pol}$ and the $d\Gamma/dq^2$ distribution. The procedure assumes a q^2 dependence of the form factors from either the ISGW model [18] or from a fit[2] of the $d\Gamma/dq^2$ distri-

[2]Here care has to be taken, since it is assumed in this fit that $R_1 = R_2 = 1$ and this

bution by Neubert [3]. The form factor ratios extracted in this procedure have errors of about 25% and 60% respectively.

The analysis presented in this dissertation will provide a more precise measurement of the form factor ratios and the form factor slope using constraints derived from the predictions of HQET. This more precise measurement is possible, even though there are more free parameters, because the available data set is larger now and full use is made of the information in the 4-dimensional distribution.

1.5 Plan of this dissertation

In chapter 2 the theory behind the measurement of the form factors is discussed. This includes the definition of the form factors, the predictions of HQS, and the definition of the parameters measured in this dissertation. The data set used, and a brief discussion of the CLEO II experiment, are presented in chapter 3. The event selection is discussed in chapter 4. Here the backgrounds in the analysis are also discussed. Chapter 5 discusses the experimental technique, including the fitting method and the calculation of the kinematic variables, and the Cramér-Rao theorem is also discussed. One important aspect of this analysis is the experimental acceptance, in particular the dependance on $\cos\theta_\ell$. The Cramér-Rao theorem provides a tool for understanding the effects of acceptance. In chapter 6 the systematic uncertainties on the measurement are discussed. The most important sources of systematic errors are the finite volume sizes used in the fit and the contamination from $B \to D^{**}\ell\nu$. Chapter 7 presents the results of the form factor fits and discusses the implications.

In Part II a project that is not related to the form factor measurement, is presented. This is the implementation of a new track fitter for CLEO II. The goal is to provide a track fitter that correctly handles the effects of multiple scattering and energy loss of the particles. The new fitter gives substantial improvemements in the track parameters at low momentum, as well as much more reliable track error matrices.

could lead to a bias in the extraction of the form factor ratios.

Chapter 2

Dynamics, form factors and heavy quark symmetry

In this chapter the invariant form factors are defined and their relationship with the helicity amplitudes is discussed. A detailed derivation of the full matrix element and the decay distribution is given in Section 2.3. In this analysis the lepton masses are set to zero. The effect of non-zero lepton masses are discussed in detail in Reference [19]. For an electron or muon in the final state the effects of non-zero masses are found to be very small[1].

2.1 Kinematics of semileptonic decays

Assuming that the D^{*+} is a narrow resonance, the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$, $D^{*+} \to D^0\pi^+$ is fully parametrized using four variables. This is easily seen by counting the number of parameters and constraints; there are $4 \times 4 = 16$ parameters, 4 constraints from 4-momentum conservation, 5 mass constraints (D^0,π^+,ℓ,ν,D^*) and three overall rotational angles. This leaves 4 variables to describe the decay. A convenient set of variables is: q^2, $\cos\theta_\ell$, $\cos\theta_V$, and χ where the angles are defined in Figure 2.1. The quantity $q^2 = (p_B - p_{D^*})^2$ is the 4-momentum transfer squared between the $\bar{B}^0$ and D^{*+}, and can also be described as the mass squared of the virtual W^-. The range of q^2 is 0 to $(M_B - M_{D^*})^2 = 10.7\ \mathrm{GeV}^2/\mathrm{c}^4$. The notation $q^2 = t = M_B^2 y$ is also used. The 4-velocity transfer $w = v_B \cdot v_{D^*}$, where $v_B = p_B/M_B$ and $v_{D^*} = p_{D^*}/M_{D^*}$ are the 4-velocities of the $\bar{B}^0$ and the D^{*+}, is simply related

[1]Terms which vanishes in the zero mass limit are suppressed by m_ℓ^2/q^2. For $m_\ell^2 = m_\mu^2 = 0.01\ \mathrm{GeV}^2/\mathrm{c}^4$ this is a a very minute effect.

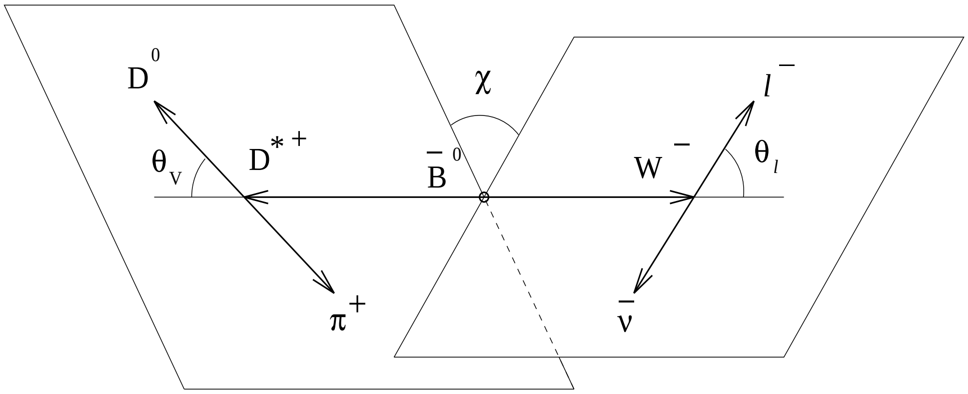

Figure 2.1: The angles used to describe the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ where $D^{*+} \to D^0\pi^+$. The B^0 is shown in its rest frame and then the decay of the W^- and the D^{*+} are shown in their rest frames.

to q^2 by

$$w = \frac{{M_B}^2 + {M_{D^*}}^2 - q^2}{2M_B M_{D^*}} = \frac{E_{D^*}}{M_{D^*}}, \tag{2.1}$$

where E_{D^*} is the energy of the D^* in the B rest frame. It is also clear from the formula above that this quantity is just the relativistic γ of the D^{*+} in the $\bar{B}^0$ rest frame. It is unfortunate that sometimes y is also used for this quantity.

In the rest frame of the virtual W^- the decay angle, θ_ℓ, of the lepton can be measured. For fixed q^2 the lepton energy E_ℓ is linearly related to $\cos\theta_\ell$ according to

$$E_\ell = E_\ell^{\rm min} + \frac{1}{2}(1+\cos\theta_\ell)\left[E_\ell^{\rm max} - E_\ell^{\rm min}\right] \tag{2.2}$$

where

$$E_\ell^{\rm max} = \frac{1}{2M_B}\left[\frac{1}{2}(M_B^2 - M_{D^*}^2 + q^2) + M_B P_{D^*}\right], \tag{2.3}$$

$$E_\ell^{\rm min} = \frac{1}{2M_B}\left[\frac{1}{2}(M_B^2 - M_{D^*}^2 + q^2) - M_B P_{D^*}\right], \tag{2.4}$$

$$P_{D^*} = \sqrt{\frac{(M_B^2 + M_{D^*}^2 - q^2)^2}{4M_B^2} - M_{D^*}^2}. \tag{2.5}$$

The quantities P_{D^*}, $E_\ell^{\rm max}$, and $E_\ell^{\rm min}$ are measured in the $\bar{B}^0$ rest frame where the above formulas hold.

In the rest frame of the D^{*+}, θ_V is the decay angle of the D^0. The angle χ measured in the $\bar{B}^0$ rest frame between the decay planes of the W^- and the D^{*+} is the last variable needed to fully specify the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$, $D^{*+} \to D^0\pi^+$. This variable is defined to be the angle measured from the half plane with the lepton to the half plane with the D^0. Note that in Reference [20] the authors define $\phi^* = \pi - \chi$. There is no standard way of defining $\cos\theta_\ell$ or $\cos\theta_V$, so comparison between different sources requires some care.

2.2 Form factors and helicity amplitudes

Here the invariant form factors governing the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ are defined. There is not a unique way of doing this, and indeed throughout this dissertation several different ways of writing down the invariant form factors will be used depending on what is convenient at the moment. The form used below is common and is used in Reference [19]. The object is to parametrize the weak-hadronic current for the transition $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$. In this transition a b-quark is destroyed and a c-quark is created. This is described by the operator

$$J^\mu = \bar{c}\gamma^\mu(1-\gamma^5)b \tag{2.6}$$

acting to the right on the initial state, the B meson, and to the left of the final state, the D^{*+} meson, giving the weak hadronic current. The current is then parametrized by a set of form factors,

$$\begin{aligned} H^\mu = \langle D^{*+}(k,\varepsilon)|\bar{c}\gamma^\mu(1-\gamma_5)b|\bar{B}^0(p)\rangle = \\ \frac{2i\epsilon^{\mu\nu\alpha\beta}}{M_B+M_{D^*}}\varepsilon^*_\nu k_\alpha p_\beta V(q^2) - (M_B+M_{D^*})\varepsilon^{*\mu}A_1(q^2) + \\ \frac{\varepsilon^*\cdot p}{M_B+M_{D^*}}(p+k)^\mu A_2(q^2) - \frac{\varepsilon^*\cdot p}{M_B+M_{D^*}}(p-k)^\mu A_3(q^2). \end{aligned} \tag{2.7}$$

This is the most general form allowed under the requirement that the weak hadronic current should transform as a vector under Lorentz transformations and that it has to be linear in ε^*, the polarization of the D^{*+}. The invariant form factors can only be functions of the available Lorentz scalars in the process. Here, there is only one such scalar, which can be chosen to be q^2 or w.

The hadronic current, H^μ, is conveniently written in the form

$$H^\mu = \varepsilon^*_\nu T^{\nu\mu} \tag{2.8}$$

where

$$T^{\nu\mu} = \frac{2i\epsilon^{\mu\nu\alpha\beta}}{M_B + M_{D^*}} k_\alpha p_\beta V(q^2) - (M_B + M_{D^*}) g^{\mu\nu} A_1(q^2) + \frac{p^\nu}{M_B + M_{D^*}} (p+k)^\mu A_2(q^2) - \frac{p^\nu}{M_B + M_{D^*}} (p-k)^\mu A_3(q^2). \quad (2.9)$$

$T^{\mu\nu}$ can be interpreted as the vertex factor for the D^*BW vertex, see Figure 2.2. For the polarization ε^ν of the D^* and ε_W^μ of the W the amplitude

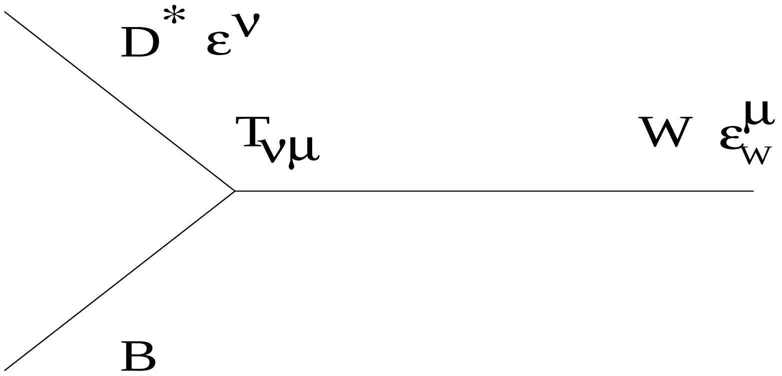

Figure 2.2: The vertex function for D^*BW.

for the transition $B \to D^*W$ is given by

$$A = \varepsilon^{*\nu} \varepsilon_W^{*\mu} T_{\nu\mu}. \quad (2.10)$$

It is now convenient to chose a basis for the polarization vectors. The basis vectors chosen are the helicity eigenstates and are labeled by the helicity, $\lambda = (t, -, 0, +)$. The D^* polarization vectors in the B rest frame are given by

$$\begin{aligned} \varepsilon_+^\mu &= \frac{1}{\sqrt{2}}(0, -1, -i, 0), \\ \varepsilon_-^\mu &= \frac{1}{\sqrt{2}}(0, 1, -i, 0), \\ \varepsilon_0^\mu &= \frac{1}{M_{D^*}}(P_{D^*}, 0, 0, E_{D^*}) \end{aligned} \quad (2.11)$$

where the D^* is moving along the z-direction in the B rest frame. The choice of the x- and y-axis orientation around the z-axis is arbitrary. The

D^* polarization vectors in the D^* rest frame are given by

$$\begin{aligned}
\varepsilon^{\mu}_{+} &= \frac{1}{\sqrt{2}}(0,-1,-i,0), \\
\varepsilon^{\mu}_{-} &= \frac{1}{\sqrt{2}}(0,1,-i,0), \\
\varepsilon^{\mu}_{0} &= (0,0,0,1).
\end{aligned} \tag{2.12}$$

The W polarization vectors in the B rest frame are given by

$$\begin{aligned}
\varepsilon^{\mu}_{W_+} &= \frac{1}{\sqrt{2}}(0,1,-i,0), \\
\varepsilon^{\mu}_{W_-} &= \frac{1}{\sqrt{2}}(0,-1,-i,0), \\
\varepsilon^{\mu}_{W_0} &= \frac{1}{\sqrt{q^2}}(P_{D^*},0,0,-q_0), \\
\varepsilon^{\mu}_{W_t} &= \frac{1}{\sqrt{q^2}}(q_0,0,0,-P_{D^*})
\end{aligned} \tag{2.13}$$

where $q_0 = M_B - E_{D^*}$. The W polarization vectors in the W rest frame are given by

$$\begin{aligned}
\varepsilon^{\mu}_{W_+} &= \frac{1}{\sqrt{2}}(0,1,-i,0), \\
\varepsilon^{\mu}_{W_-} &= \frac{1}{\sqrt{2}}(0,-1,-i,0), \\
\varepsilon^{\mu}_{W_0} &= (0,0,0,-1), \\
\varepsilon^{\mu}_{W_t} &= (1,0,0,0).
\end{aligned} \tag{2.14}$$

Using these basis vectors, the helicity amplitudes are defined by projecting out the amplitudes from the vertex function

$$H_{\lambda_{D^*}\lambda_W} = \varepsilon^{*\nu}_{\lambda_{D^*}} \varepsilon^{*\mu}_{W_{\lambda_W}} T_{\nu\mu}. \tag{2.15}$$

Of these 12 helicity amplitudes only 4 are nonzero

$$H_{+} \equiv H_{++} = -(M_B + M_{D^*})A_1(q^2) + 2\frac{P_{D^*}M_B}{M_B + M_{D^*}}V(q^2), \tag{2.16}$$

$$H_{-} \equiv H_{--} = -(M_B + M_{D^*})A_1(q^2) - 2\frac{P_{D^*}M_B}{M_B + M_{D^*}}V(q^2), \tag{2.17}$$

$$H_0 \equiv H_{00} = -\frac{1}{2M_{D^*}\sqrt{q^2}}(A_1(q^2)(M_B + M_{D^*})(M_B^2 - M_{D^*}^2 - q^2) -$$

$$4\frac{M_B^2 P_{D^*}^2}{M_B + M_{D^*}} A_2(q^2)), \tag{2.18}$$

$$\begin{aligned} H_t \equiv H_{0t} &= -\frac{M_B P_{D^*}}{M_{D^*}\sqrt{q^2}}(A_1(q^2)(M_B + M_{D^*}) - \\ &\frac{1}{2}(M_B^2 - M_{D^*}^2 + q^2)\frac{A_2(q^2)}{(M_B + M_{D^*})} + q^2\frac{A_3(q^2)}{M_B + M_{D^*}}). \end{aligned} \tag{2.19}$$

That these are the only nonzero elements is easily seen since the helicity has to be conserved in the decay.

2.3 Differential decay rate

In this section the differential decay rate is derived. There are several ways of doing this, the conceptually simplest method is the helicity formalism [21]. This method has the danger of depending on conventions, in particular, the conventions for defining the helicity amplitudes, Eqs. 2.16- 2.19. The derivation presented here follows closely that of Reference [19]. The advantage of this method is that it is free of conventions.

2.3.1 Notation and conventions

In the calculation below the following notation and conventions are used.

$$\epsilon_{0123} = -\epsilon^{0123} = 1 \tag{2.20}$$

$$\eta_t = 1 \qquad \eta_{+,0,-} = -1 \tag{2.21}$$

$$g_{mn} = \mathrm{diag}(1, -1, -1, -1) \qquad m, n = (t, +, -, 0) \tag{2.22}$$

$$p^\mu = (E, \vec{P}) \tag{2.23}$$

The polarization vectors defined in Section 2.2 satisfies the following completeness relations

$$\sum_{m,m'} \varepsilon_m^\alpha \varepsilon_{m'}^{*\beta} \delta_{mm'} = -g^{\alpha\beta} + \frac{k^\alpha k^\beta}{M_{D^*}^2} = S_{M_{D^*}}^{\alpha\beta}, \tag{2.24}$$

$$\sum_{m,m'} \varepsilon_{W_m}^\mu \varepsilon_{W_{m'}}^{*\nu} g_{mm'} = g^{\mu\nu}. \tag{2.25}$$

These relations will be used below in the derivation of the differential decay rate. Latin indices $(\alpha, \beta, \mu, \nu)$ are used for four vector indices, roman indices (m, n) are used for helicity indices.

2.3.2 Lorentz invariant phase space

Here the Lorentz invariant phase space is evaluated. This is a four-body decay;

$$d\phi_4(P;p_1,p_2,p_3,p_4) = \delta^4(P-\sum_{i=1}^{4}p_i)\prod_{i=1}^{4}\frac{d^3p_i}{(2\pi)^3 2E_i} \tag{2.26}$$

is needed. Using the recursive relation

$$d\phi_n(P;p_1,p_2...p_n) = d\phi_j(q;p_1...p_j)d\phi_{n-j+1}(P;q,p_{j+1}...p_n)(2\pi)^3dq^2 \tag{2.27}$$

twice the four body phase space can be written

$$\begin{aligned} &d\phi_4(P;p_l,p_\nu,p_D,p_\pi) = \\ &\quad (2\pi)^6 d\phi_2(P;k,q)d\phi_2(k;p_D,p_\pi)d\phi_2(q;p_l,p_\nu)dq^2dk^2. \end{aligned} \tag{2.28}$$

The two body phase space is easily evaluated from the definition and is in the rest frame of the parent found to be

$$d\phi_2(P,k,q) = \frac{|\vec{q}|}{P_0}\frac{d\Omega_q}{4(2\pi)^6}. \tag{2.29}$$

Putting this together the final expression for the four-body phase space is

$$d\phi_4(P;p_l,p_\nu,p_D,p_\pi) = \frac{1}{32(2\pi)^{10}}\frac{|\vec{P}_{D^*}|}{M_B}\frac{|\vec{P}_{D^0}|}{M_{D^*}}d\cos\theta_V d\chi d\cos\theta_\ell dq^2dk^2. \tag{2.30}$$

The three-body phase is given by

$$d\phi_3(P;p_l,p_\nu,k) = \frac{1}{16(2\pi)^7}\frac{|\vec{P}_{D^*}|}{M_B}d\cos\theta_\ell dq^2. \tag{2.31}$$

All 3-momenta are measured in their parent's rest frame.

2.3.3 Calculation of the decay rate

The differential decay rate can now be written down. As a first step consider the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$. This is done so that the decay $D^{*+}\to D^0\pi^+$ can be properly normalized. Using standard Feynman rules for electroweak theory the amplitude for this process is given by

$$\mathcal{A}(\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}) =$$

$$\langle D^{*+}(k,\varepsilon)|\bar{c}\frac{-ig}{2\sqrt{2}}V_{cb}\gamma_\mu(1-\gamma_5)b|\bar{B}^0(p)\rangle \times \tag{2.32}$$

$$P^{\mu\nu}(q)\bar{u}_\ell\frac{-ig}{2\sqrt{2}}\gamma_\nu(1-\gamma_5)v_\nu \tag{2.33}$$

where $P^{\mu\nu}(q)$ is the W propagator and q is the 4-momentum of the W.

$$P^{\mu\nu}(q) = \frac{i(-g^{\mu\nu} + q^\mu q^\nu / M_W^2)}{q^2 - M_W^2} \simeq i\frac{g^{\mu\nu}}{M_W^2}, \tag{2.34}$$

where the approximation that the momentum transfer is much smaller than the W mass is used. This gives

$$\mathcal{A}(\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}) = -i\frac{G_F}{\sqrt{2}} V_{cb} L^\mu H_\mu, \tag{2.35}$$

where

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2}, \tag{2.36}$$

$$L^\mu = \bar{u}_\ell \gamma^\mu (1-\gamma_5) v_\nu, \tag{2.37}$$

and

$$H_\mu = \langle D^{*+}(k,\varepsilon)|\bar{c}\gamma_\mu(1-\gamma_5)b|\bar{B}^0(p)\rangle. \tag{2.38}$$

In Section 2.2 the form of the hadronic current was discussed and the form factors defined. H_μ was written as

$$H_\mu = \varepsilon^{*\nu} T_{\mu\nu} \tag{2.39}$$

where $T_{\mu\nu}$ is interpreted as the vertex factor for the BD^*W vertex.

Now the matrix element squared for the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ can be written as

$$|\mathcal{M}|^2 = \frac{G_F^2 |V_{cb}|^2}{2} L^{\mu\nu} H_{\mu\nu}, \tag{2.40}$$

where

$$L^{\mu\nu} = L^\mu L^{*\nu} \tag{2.41}$$

and

$$H_{\mu\nu} = H_\mu H_\nu^* = \varepsilon^{*\alpha} T_{\mu\alpha} \varepsilon^\beta T_{\nu\beta}^*. \tag{2.42}$$

The total rate is obtained by summing over the final state spins, i.e., summing over the polarizations of the D^*. After summing over the three possible polarization vectors for the D^* the hadron tensor reads

$$H_{\mu\nu} = T_{\mu\alpha}(T_{\nu\beta}^*)(-g^{\alpha\beta} + \frac{k^\alpha k^\beta}{M_{D^*}^2}). \tag{2.43}$$

Here the completeness of the polarization vectors and the fact that T is independent of the helicity have been used. The reason for doing the case of $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ is so that the decay tensor for $D^{*+} \to D^0\pi^+$ can be normalized. To calculate the rate for $\bar{B}^0 \to D^0\pi^+\ell^-\bar{\nu}$ the hadronic tensor needs to be modified so that it propagates the D^{*+} and then correctly decays it to $D^0\pi^+$.

Using Lorentz invariance it is easily found that the amplitude for D^{*+} to decay to $D^0\pi^+$ is proportional to $\varepsilon^\mu p_{D\mu}$ where p_D is the D^0 momentum. By simply removing the external factor ε^μ it is found that the vertex factor is $Cp_{D\mu}$, where C is an arbitrary constant to be determined below. Putting these things together the hadronic tensor is

$$H_{\mu\nu} = |C|^2 T_{\mu\alpha} T^*_{\nu\beta} S^{\alpha\alpha'}_{M_{D^*}} S^{\beta\beta'}_{M_{D^*}} p_{D\alpha'} p_{D\beta'} \delta(k^2 - M^2_{D^*}) \tag{2.44}$$

where $S^{\alpha\beta}_{M_{D^*}} = -g^{\alpha\beta} + \frac{k^\alpha k^\beta}{M_2^2}$ is the spin-1 propagator for the D^*. The delta function is put in by hand to make the D^* on mass shell. Defining the strong decay tensor $Z^{\alpha\beta}$ by

$$H_{\mu\nu} = T_{\mu\alpha} T^*_{\nu\beta} Z^{\alpha\beta} \tag{2.45}$$

where the normalization of $Z^{\alpha\beta}$ is given by

$$\int dk^2 \int d\chi \int d\cos\theta_V Z^{\alpha\beta} = 2(2\pi)^3 \frac{M_{D^*}}{P_{D^0}} S^{\alpha\beta}_{M_{D^*}}. \tag{2.46}$$

This is just saying that when integrated over all final state configurations of the $D^{*+} \to D^0\pi^+$ decay the same rate as for the decay $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ is obtained. Explicit calculation gives that

$$Z^{\alpha\beta} = \frac{3(2\pi)^2 M_{D^*}}{P^3_{D^0}} S^{\alpha\alpha'}_{M_{D^*}} S^{\beta\beta'}_{M_{D^*}} p_{D\alpha'} p_{D\beta'} \delta(k^2 - M^2_{D^*}). \tag{2.47}$$

Putting things together the matrix element for the decay $\bar{B}^0 \to D^0\pi^+\ell^-\bar{\nu}$ can be written

$$\begin{aligned} |\mathcal{M}|^2 &= \frac{G_F^2 |V_{cb}|^2}{2} L^{\mu\nu} T_{\mu\alpha} T^*_{\nu\beta} Z^{\alpha\beta} \\ &= \frac{6\pi^2 G_F^2 |V_{cb}|^2 M_{D^*}}{P^3_{D^0}} L_{\mu'\nu'} g^{\mu'\mu} g^{\nu'\nu} T_{\mu\alpha} T^*_{\nu\beta} S^{\alpha\alpha'}_{M_{D^*}} S^{\beta\beta'}_{M_{D^*}} p_{D\alpha'} p_{D\beta'} \end{aligned} \tag{2.48}$$

To proceed and get the equations in a form that is easy to evaluate it is convenient to use the completeness relations of the polarization vectors.

$$|\mathcal{M}|^2 = \frac{6\pi^2 G_F^2 |V_{cb}|^2 M_{D^*}}{P^3_{D^0}} \sum_{m,m'} \sum_{\tilde{m},\tilde{m}'} \sum_{n,n'} \sum_{\tilde{n},\tilde{n}'} L_{\mu'\nu'} \varepsilon^{\mu'}_{W_m} \varepsilon^{*\nu'}_{W_{\tilde{m}}} g_{mm'} g_{\tilde{m}\tilde{m}'}$$

$$\varepsilon^{*\mu}_{W_{m'}} T_{\mu\alpha} \varepsilon^{*\alpha}_{n} \varepsilon^{\nu}_{W_{\tilde{m}'}} T^{*}_{\nu\beta} \varepsilon^{\beta}_{\tilde{n}} \varepsilon^{\alpha'}_{\tilde{n}} p_{D\alpha'} \varepsilon^{*\beta'}_{n'} p_{D\beta'} \delta_{mm'} \delta_{\tilde{m}\tilde{m}'}$$

$$= \frac{6\pi^2 G_F^2 |V_{cb}|^2 M_{D^*}}{P^3_{D^0}} \sum_m \sum_{\tilde{m}} \sum_n \sum_{\tilde{n}} L_{\mu'\nu'} \varepsilon^{\mu}_{W_m} \varepsilon^{*\nu}_{W_{\tilde{m}}} \eta_m \eta_{\tilde{m}}$$

$$\varepsilon^{*\mu}_{W_m} T_{\mu\alpha} \varepsilon^{*\alpha}_{n} \varepsilon^{\nu}_{W_{\tilde{m}}} T^{*}_{\nu\beta} \varepsilon^{\beta}_{\tilde{n}} \varepsilon^{\alpha'}_{n} p_{D\alpha'} \varepsilon^{*\beta'}_{\tilde{n}} p_{D\beta'}. \quad (2.49)$$

The advantage of writing the matrix element in this form is that the expression factorizes into three pieces that can be evaluated in three different frames.

$$L^{mn} = L_{\mu\nu} \varepsilon^{\mu}_{W_m} \varepsilon^{*\nu}_{W_n} \eta_m \eta_n \quad (2.50)$$

which can be evaluated in the W rest frame. Using

$$L^{\mu\nu} = 8(p^{\mu}_{\ell} p^{\nu}_{\nu} + p^{\nu}_{\ell} p^{\mu}_{\nu} - \frac{q^2}{2} g^{\mu\nu} + i\epsilon^{\mu\nu\alpha\beta} p_{\ell\alpha} p_{\nu\beta}) \quad (2.51)$$

and the explicit forms

$$\begin{aligned} p^{\mu}_{\ell} &= (P_l, P_l \sin\theta_\ell \cos\chi, P_l \sin\theta_\ell \sin\chi, -P_l \cos\theta_\ell), && (2.52) \\ p^{\mu}_{\nu} &= (P_l, -P_l \sin\theta_\ell \cos\chi, -P_l \sin\theta_\ell \sin\chi, P_l \cos\theta_\ell) && (2.53) \end{aligned}$$

in the W rest frame for the lepton and neutrino four vectors. $P_l = \sqrt{q^2}/2$ is the maginitude of the lepton momentum in the W rest frame. The following nine non-zero elements of L^{mn} are found

$$\begin{aligned} L^{++} &= 8P_l^2 (\cos\theta_\ell - 1)^2, && (2.54) \\ L^{+0} &= 8\sqrt{2} P_l^2 \sin\theta_\ell (\cos\theta_\ell - 1)(\cos\chi - i\sin\chi), && (2.55) \\ L^{+-} &= 8P_l^2 \sin^2\theta_\ell (\cos\chi - i\sin\chi)^2, && (2.56) \\ L^{0+} &= -8\sqrt{2} P_l^2 \sin\theta_\ell (\cos\theta_\ell - 1)(\cos\chi + i\sin\chi), && (2.57) \\ L^{00} &= 16 P_l^2 \sin^2\theta_\ell, && (2.58) \\ L^{0-} &= 8\sqrt{2} P_l^2 \sin\theta_\ell (\cos\theta_\ell + 1)(\cos\chi - i\sin\chi), && (2.59) \\ L^{-+} &= 8P_l^2 \sin^2\theta_\ell (\cos\chi + i\sin\chi)^2, && (2.60) \\ L^{-0} &= 8\sqrt{2} P_l^2 \sin\theta_\ell (\cos\theta_\ell + 1)(\cos\chi + i\sin\chi), && (2.61) \\ L^{--} &= 8P_l^2 (1 + \cos\theta_\ell)^2. && (2.62) \end{aligned}$$

In the rest frame of the D^*

$$K^m = \varepsilon^{\mu}_{m} p_{D\mu} \quad (2.63)$$

with $p_D^\mu = (E_D, P_D \sin\theta_V, 0, P_D \cos\theta_V)$

$$\begin{aligned}
K^+ &= \frac{1}{\sqrt{2}} P_D \sin\theta_V, && (2.64)\\
K^0 &= -P_D \cos\theta_V, && (2.65)\\
K^- &= -\frac{1}{\sqrt{2}} P_D \sin\theta_V. && (2.66)
\end{aligned}$$

Finally, in the B rest frame

$$H_{mn} = \varepsilon_m^{*\alpha} \varepsilon_{W_n}^{*\mu} T_{\mu\alpha} \qquad (2.67)$$

There are only three non-vanishing elements of H_{mn} that contribute in the zero lepton mass limit

$$\begin{aligned}
H_+ &\equiv H_{++} = -(M_B + M_{D^*}) A_1(q^2) + 2\frac{P_{D^*} M_B}{M_B + M_{D^*}} V(q^2), && (2.68)\\
H_- &\equiv H_{--} = -(M_B + M_{D^*}) A_1(q^2) - 2\frac{P_{D^*} M_B}{M_B + M_{D^*}} V(q^2), && (2.69)\\
H_0 &\equiv H_{00} = -\frac{1}{2M_{D^*}\sqrt{q^2}} (A_1(q^2)(M_B + M_{D^*})(M_B^2 - M_{D^*}^2 - q^2) - \\
&\quad 4\frac{M_B^2 P_{D^*}^2}{M_B + M_{D^*}} A_2(q^2)). && (2.70)
\end{aligned}$$

Putting this together

$$\begin{aligned}
|\mathcal{M}|^2 &= \frac{G_F^2 |V_{cb}|^2}{2} L^{\mu\nu} H_{\mu\nu} \\
&= \frac{6\pi^2 G_F^2 |V_{cb}|^2 M_{D^*}}{P_{D^0}^3} H_{\tilde{m}m} H_{n\tilde{n}}^* L^{mn} K^{\tilde{m}} K^{\tilde{n}} \delta(k^2 - M_{D^*}^2) \\
&= \frac{24\pi^2 G_F^2 |V_{cb}|^2 M_{D^*} P_l^2}{P_{D^0}} \delta(k^2 - M_{D^*}^2) \\
&\quad (|H_+|^2 (1-\cos\theta_\ell)^2 \sin^2\theta_V + |H_-|^2 (1+\cos\theta_V)^2 \sin^2\theta_V \\
&\quad +4|H_0|^2 \sin^2\theta_\ell \cos^2\theta_V - 2H_+ H_- \sin^2\theta_\ell \sin^2\theta_V \cos(2\chi) \\
&\quad -4H_+ H_0 \sin\theta_\ell (1-\cos\theta_V) \sin\theta_V \cos\theta_V \cos\chi \\
&\quad +4H_- H_0 \sin\theta_\ell (1+\cos\theta_V) \sin\theta_V \cos\theta_V \cos\chi). && (2.71)
\end{aligned}$$

The decay rate is finally given by

$$d\Gamma = \frac{(2\pi)^4}{2M_B} |\mathcal{M}|^2 d\phi_4(P; p_l, p_\nu, p_D, p_\pi) \qquad (2.72)$$

which gives

$$\frac{d\Gamma}{dq^2 d\cos\theta_\ell d\cos\theta_V d\chi} = \frac{3G_F^2|V_{cb}|^2 P_{D^*} q^2}{8(4\pi)^4 M_B^2} B(D^{*+} \to D^0\pi^+)$$
$$(|H_+|^2(1-\cos\theta_\ell)^2 \sin^2\theta_V + |H_-|^2(1+\cos\theta_\ell)^2 \sin^2\theta_V$$
$$+4|H_0|^2 \sin^2\theta_\ell \cos^2\theta_V - 2H_+H_- \sin^2\theta_\ell \sin^2\theta_V \cos(2\chi)$$
$$-4H_+H_0 \sin\theta_\ell(1-\cos\theta_\ell) \sin\theta_V \cos\theta_V \cos\chi$$
$$+4H_-H_0 \sin\theta_\ell(1+\cos\theta_\ell) \sin\theta_V \cos\theta_V \cos\chi) \quad (2.73)$$

This is the final form of the differential decay rate. Here the branching fraction for $D^{*+} \to D^0\pi^+$ has been inserted since the calculation assumed a 100% branching fraction for this decay.

2.3.4 Properties of the semileptonic decay

Before looking at the predictions of quark models and heavy quark symmetry some model independent predictions of these distributions will be discussed.

The first observation is that the W couples to particles with left-handed chirality and antiparticles with right-handed chirality. In the limit of relativistic velocities the chirality projection operator becomes the same as the helicity projection operator. This means that the produced charm quark will prefer a negative helicity. This gets inherited by the produced D^* and means that there is a larger probability for the produced D^* to have helicity -1 than $+1$. This is reflected in the helicity amplitude by the expectation that $|H_-| > |H_+|$ for a given q^2. From this argument it is impossible to compare the helicity zero amplitude, H_0, to $H_\pm$. Since the helicity amplitudes are interpreted as the amplitudes for either the D^* or the virtual W, since the B is a scalar, also the helicity of the W is more likely to be negative than positive. Since again, when the W decays it couples to left-handed particles and right-handed antiparticles the lepton prefers to be forward in W rest frame. This gives rise to a hard lepton spectra, as compared to the semileptonic decay of D mesons in which the lepton likes to be backward in the W rest frame. This effect is obvious when looking at the Dalitz plot, q^2 vs. E_ℓ, where the events prefer high lepton energy, as shown in Figure 2.3. In Figure 2.4 the different contributions to the rate from the helicity amplitudes and their interference are shown for the HQS predictions. The amplitude for the decay to have a negative helicity W is significantly larger than for positive helicity. HQS predicts, in the limit of infinitely heavy b and c quark masses, the relative rate of the three different helicities. It does not predict the q^2 dependence; a linear form of $h_{A_1}(w)$ is assumed.

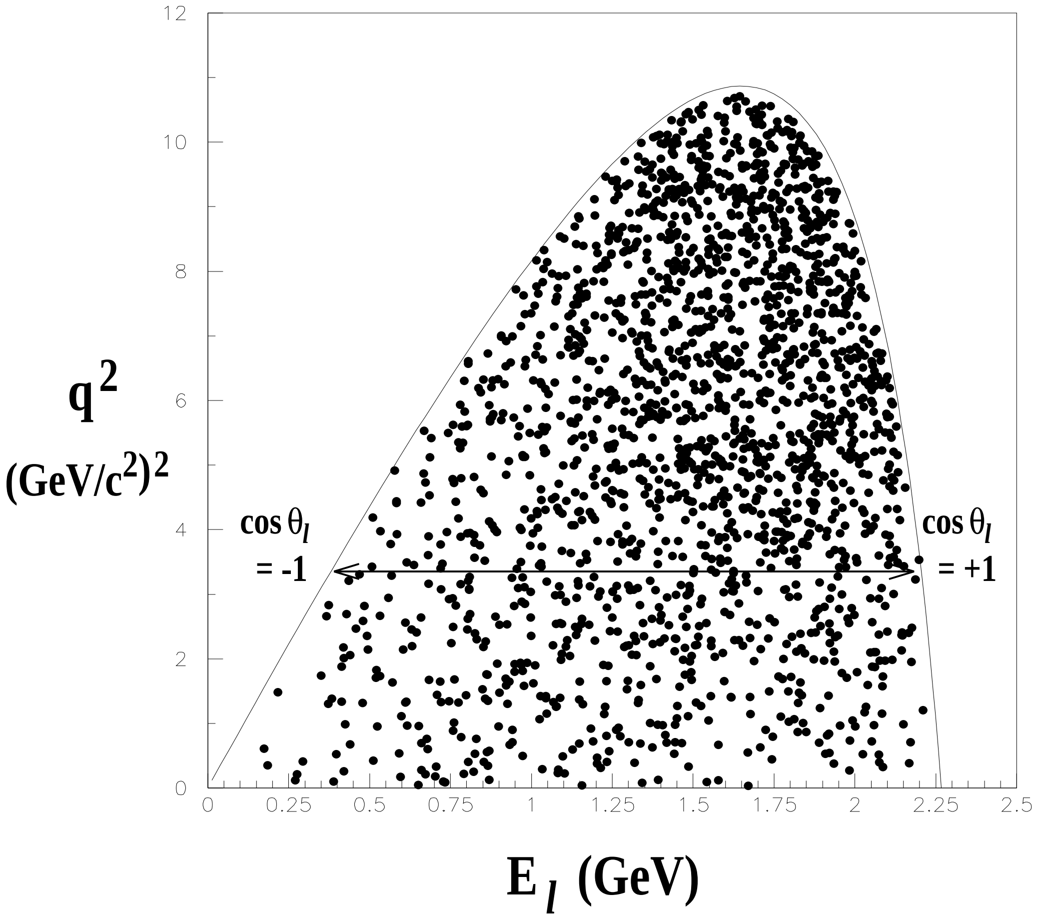

Figure 2.3: Dalitz plot, q^2 vs. E_ℓ, for $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$. The events are generated according to the predictions of HQET, $R_1 = 1$ and $R_2 = 1$ with the linear form $H_{A_1}(w) = H_{A_1}(1)[1 - \rho^2_{A_1}(w-1)]$ with $\rho^2_{A_1} = 1$. For a given q^2 the lepton energy, E_ℓ, varies linearly with $\cos\theta_\ell$, from $\cos\theta_\ell = -1$ at $E_\ell^{\min}$ to $\cos\theta_\ell = +1$ at $E_\ell^{\max}$.

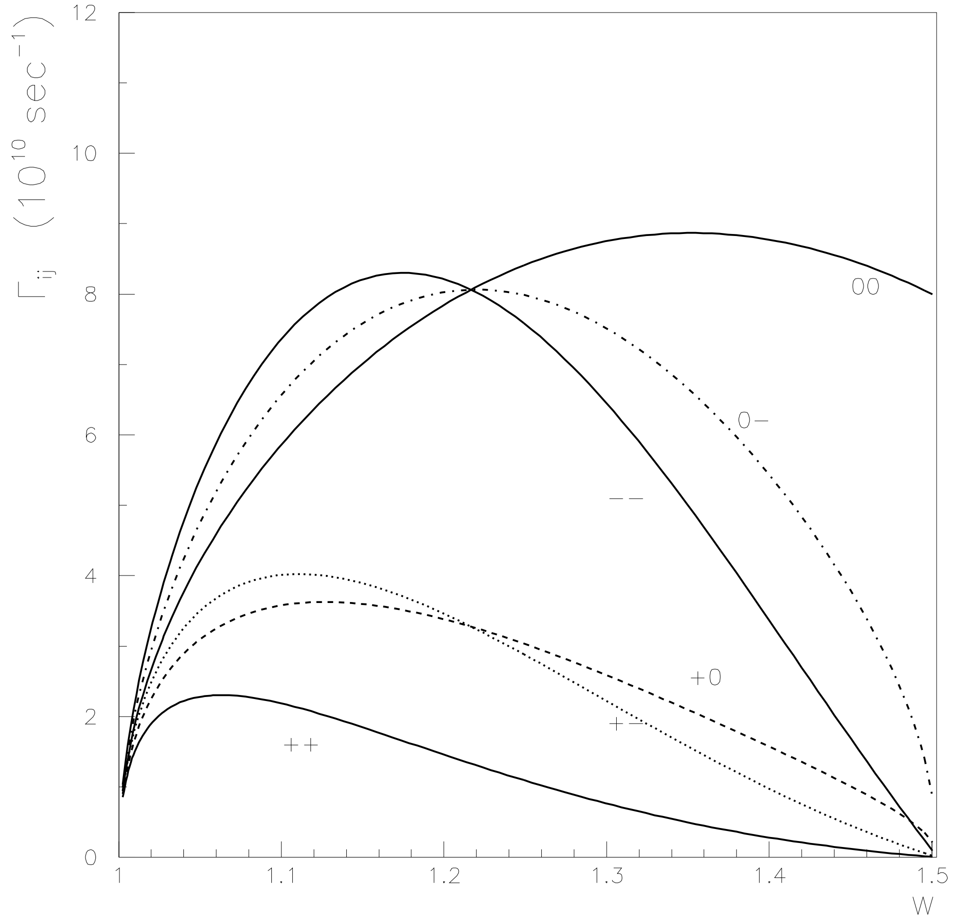

Figure 2.4: The contribution to the decay rate from the different helicity amplitudes as a function of w. The quantities Γ_{ij} are defined by $\Gamma_{ij} = \dfrac{G_F^2|V_{cb}|^2 P_{D^*} q^2}{96\pi^3} H_i H_j$, $(i, j = \pm, 0)$. The solid lines are the contributions from the diagonal elements $(i = j)$. Note that the contribution to the rate from $|H_-|^2$ is much bigger than the contribution from $|H_+|^2$. This effect is due to of the parity violation in the decay. The total rate is $\Gamma = \int dq^2(\Gamma_{++} + \Gamma_{00} + \Gamma_{--})$. Although the off diagonal terms, Γ_{ij} $i \neq j$, do not contribute to the total decay rate, the effect of these terms can be seen in scatter plots of the kinematic variables.

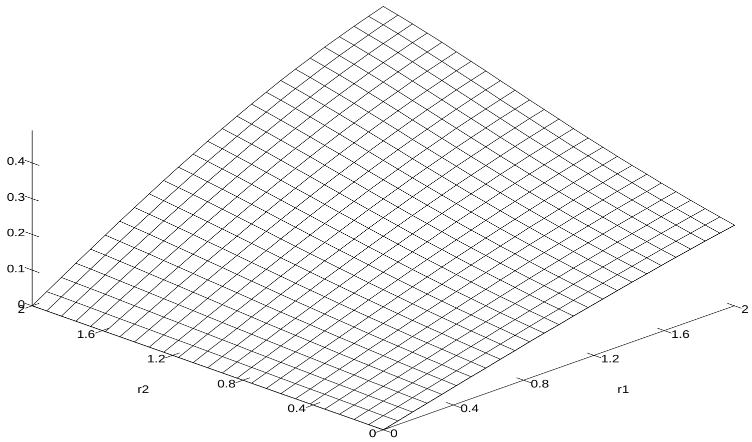

Figure 2.5: The forward backward asymmetry as a function of the two form factor ratios R_1 and R_2. The form factor slope is fixed to $\rho^2_{A_1} = 1.0$.

The distribution in $\cos\theta_V$ is easily understood in the two extreme cases of q^2. At $q^2 = q^2_{\rm max}$ the D^* is produced at rest and the helicity states of the D^* are equally populated giving rise to a uniform distribution in $\cos\theta_V$. For $q^2 = 0$, the maximum recoil configuration, the lepton and the neutrino momenta are parallel, forcing the helicity of the W, and the D^*, to be zero. This gives a $\cos^2\theta_V$ distribution in this limit.

It is expected that the q^2 distribution has a pole in the nearest bc resonance with the right spin-parity. This would be the B_c^* resonance with a mass around 6.7 GeV/c^2. This resonance is far outside the Dalitz plot boundary but will still affect the distribution in the Dalitz plot by enhancing the amplitude at high q^2. Another way of arguing this is that at the $q^2 = q^2_{\rm max}$ configuration in which the produced D^* is at rest the probability for the slow moving charm quark to form a meson is large while as the momentum of the charm quark increases (q^2 decreases) the probability for the charm quark to form a D^* with the spectator quark is smaller.

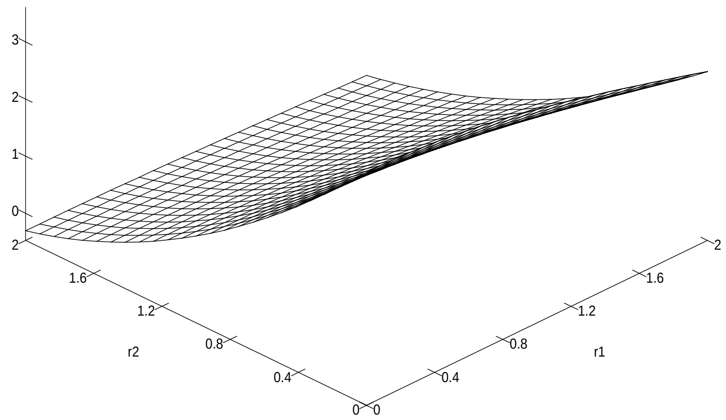

Figure 2.6: The longitudinal to transverse polarization of the D^{*+} as a function of the two form factor ratios R_1 and R_2. The form factor slope is fixed to $\rho^2_{A_1} = 1.0$.

2.4 Form factor predictions

There are several models that are used to to predict the rates and distributions of semileptonic decays. Here, the four models ISGW [18], KS [19], WSB [22], and ISGW2 [23] are discussed. The ISGW2 model is an update of the original ISGW model, but there are quite large modifications. All these models make predictions—and assumptions—about the form factors. In particular the models assume a form of the form factors and predict the normalization. The models of ISGW, KS, and WSB were pre-HQS models while the ISGW2 model incorporates the constraints from HQS.

In the ISGW model the form factors are assumed to be of the form

$$F(q^2) = Ce^{-a(q^2_{\mathrm{max}} - q^2)} \tag{2.74}$$

where a is a constant that describes how fast the form factor is falling as q^2 goes to zero. This constant is related to the size of the mesons and is assumed to be the same for all form factors in a given decay. The constant C is the normalization, which is calculated independently for all form factors in the decay. In the ISGW model this normalization is calculated at the $q^2 = q^2_{\mathrm{max}}$ configuration where the final state meson is at rest in the parent meson's rest frame. In the ISGW model this normalization is obtained from the overlap of the non relativistic quark model wave functions.

The BSW model uses a pole form for the form factors

$$F(q^2) = \frac{F(0)}{1 - q^2/M^2_{\rm pole}}. \tag{2.75}$$

The pole mass is fixed to the lowest mass resonance that has the right quantum numbers, spin-parity and quark content. In this model the normalization is done at the $q^2 = 0$ configuration.

The KS model is a variation of the BSW model in that it uses the same normalization but the form factors are of the form

$$F(q^2) = \frac{F(0)}{(1 - q^2/M^2_{\rm pole})^n} \tag{2.76}$$

where n is determined from QCD power counting rules.

In the ISGW2 model the authors have abandoned the exponential form of the form factors for a pole form. However, the pole is not expressed as a mass. Instead the authors expresse the q^2 fall-off in terms of the meson radius. The model has also been updated to include constraints from heavy quark symmetry.

2.5 $\bar{A}_{\rm FB}$ and $\bar{A}_{\rm pol}$

There are two common observables that are related to the form factors. They are the forward-backward asymmetry, $\bar{A}_{\rm FB}$, of the lepton in the W rest frame and the longitudinal to transverse polarization, $\bar{A}_{\rm pol}$ of the D^{*+}. They are defined by

$$\bar{A}_{\rm FB} = \frac{\displaystyle\int_0^1 \frac{d\Gamma}{d\cos\theta_\ell} d\cos\theta_\ell - \int_{-1}^0 \frac{d\Gamma}{d\cos\theta_\ell} d\cos\theta_\ell}{\displaystyle\int_{-1}^1 \frac{d\Gamma}{d\cos\theta_\ell} d\cos\theta_\ell}, \tag{2.77}$$

$$\begin{aligned} \bar{A}_{\rm pol} &= 2\frac{\displaystyle\int P_{D^*} q^2 |H_0(q^2)|^2 dq^2}{\displaystyle\int P_{D^*} q^2 (|H_-(q^2)|^2 + |H_+(q^2)|^2) dq^2} - 1 \\ &= 2\frac{\Gamma_L}{\Gamma_T} - 1. \end{aligned} \tag{2.78}$$

In Figure 2.5 and 2.6 the forward backward asymmetry of the lepton and the longitudinal to transverse polarization of the D^{*+} are shown as functions of R_1 and R_2 for a fixed form factor slope, $\rho^2_{A_1} = 1.0$. The forward-backward asymmetry of the lepton is mainly controlled by R_1. This is understood since R_1 controls how much bigger $|H_-|$ is compared to $|H_+|$; a big $|H_-|$ implies that there is a harder lepton energy spectrum — a larger asymmetry. The longitudinal to transverse polarization, on the other hand, is almost completely controlled by R_2. R_2 controls the size of $|H_0|$ and therefore controls $\bar{A}_{\text{pol}}$.

After integrating over all angles the decay rate formula, Eq. 2.73, reads

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2|V_{cb}|^2}{96\pi^3 M_B^2} P_{D^*} q^2 \{|H_+(q^2)|^2 + |H_-(q^2)|^2 + |H_0(q^2)|^2\}. \tag{2.79}$$

A word on conventions and notation. A statement like $|H_-| > |H_+|$ applies to the case of a meson containing a b-quark. Since the decay is CP invariant the decay distribution for a meson containing a $\bar{b}$-quark will be the same. Here the amplitude for the decay to go with a positive helicity W is bigger then for the decay to go with negative helicity. By convention decays are described for mesons containing quarks that decay, as opposed to antiquarks. In Reference [24] B is defined to be the meson containing a b-quark.

2.6 Heavy quark effective theory

It was recently realized that systems containing heavy quarks or heavy quark transitions have additional symmetries [14, 15]. The first symmetry is a flavor symmetry that is in its nature similar to the isospin symmetry for the light quarks. For the heavy quark systems the mass of the heavy quark decouples from the system since it is large compared to Λ_{QCD} or the relevant energy scale. Also, as the mass of the quarks gets large the spin decouples from the dynamics and an additional spin symmetry appears.

First the heavy quark symmetry, HQS, limit will be discussed. In this limit the heavy quark masses are taken to infinity. In the transitions that are of interest here it means that $m_b, m_c \to \infty$. It will be shown how this leads to relations between the invariant form factors defined in Section 2.2. It will also be discussed how this leads to the absolute normalization of the decay rate in the zero recoil configuration.

However, the b- and c-quark masses are not infinite and the effects of the finite masses are discussed and predictions of the symmetry breaking are presented. This is done in the framework of heavy quark effective theory, HQET, in which the QCD Lagrangian is expanded in powers of E/m_b and E/m_c, where E is the relevant energy scale of the light degrees of freedom. E is approximately equal to $\Lambda_{\rm QCD}$ for the configuration in which the final state meson is at rest in the parent meson's rest frame.

This work will not contain, in any way, a complete discussion about HQS, the discussion below will focus on results of HQS applied to $B \to D^* \ell\nu$ and $B \to D\ell\nu$. The reason for treating D and D^* together is that they belong to the same heavy quark multiplet; in the limit of infinite quark masses the D^* and the D are degenerate. Already here an effect of the finite quark masses is seen; the heavy quark spin has not completely decoupled from the dynamics and a mass difference is observed. This mass difference is expected to be proportional to $1/m_Q$. Indeed, this is verified in the mass differences of the $D^* - D$ and the $B^* - B$.

2.6.1 Heavy quark symmetry

The goal of this section is to derive the most important results of heavy quark symmetry for the semileptonic decay of a pseudoscalar to a vector meson. This includes deriving the relations between the form factors in the symmetry limit and showing that the universal form factor, the Isgur-Wise function, is absolutely normalized in the heavy quark limit for the zero-recoil configuration. The results presented here were first obtained in Reference [15].

Since the symmetry studied is a spin-flavor symmetry in the limit of infinite quark masses it is convenient to use variables that are well behaved in this limit. For this reason it is convenient to describe the dynamics in terms of the four velocity, $v^\mu = p^\mu/M$, instead of the momentum, p^μ. It is also convenient to us the normalization

$$\langle M(v')|M(v)\rangle = \frac{2P^0}{M_M}(2\pi)^3\delta^3(\vec{p}-\vec{p'}) \tag{2.80}$$

instead of the more conventional relativistic normalization

$$\langle \tilde{M}(v')|\tilde{M}(v)\rangle = 2P^0(2\pi)^3\delta^3(\vec{p}-\vec{p'}). \tag{2.81}$$

This gives

$$\langle M(v)| = M_M^{1/2}\langle \tilde{M}(v)|. \tag{2.82}$$

In the limit of infinitely heavy quarks the state of the meson, $\langle M(v)|$, is completely specified by the state of the light degrees of freedom. So with the mass independent normalization the states are independent of the mass. This makes intuitive sense since the heavy quark will act as a static source of color charge independent of the flavor.

The heavy quark field is denoted by $h_v(x)$ where v is the velocity of the heavy quark. h_v annihilates a heavy quark of velocity v. Similarly the field $\bar{h}_v$ creates a heavy quark with velocity v. These heavy quark fields are subject to an on-shell condition which is just the Dirac equation

$$\begin{aligned} \not{v} h_v &= h_v, & (2.83) \\ \bar{h}_v \not{v} &= \bar{h}_v. & (2.84) \end{aligned}$$

Now the pseudoscalar to pseudoscalar transition in the heavy quark symmetry limit will be discussed. This is, for example, the transition $\bar{B}^0 \to D^+\ell^-\bar{\nu}$. The initial pseudoscalar meson is denoted by $P(v)$ and the final state meson by $P'(v')$, where v and v' are the four-velocities of the two mesons. Based on the requirements from Lorentz invariance the weak hadronic current can be written

$$\begin{aligned} \langle P'(v')|\bar{h}_{v'}\gamma^\mu(1-\gamma^5)h_v|P(v)\rangle = \langle P'(v')|\bar{h}_{v'}\gamma^\mu h_v|P(v)\rangle &= \\ \xi^+(v\cdot v')(v+v')^\mu + \xi^-(v\cdot v')(v-v')^\mu. & \quad (2.85) \end{aligned}$$

By contracting both sides of the last equality by $(v-v')_\mu$ and using Eq. 2.84 it follows that $\xi^-(v\cdot v') = 0$ since $(v-v')_\mu \bar{h}_{v'}\gamma^\mu h_v = 0$, $(v-v')^\mu(v+v')_\mu = 0$, and $(v-v')^\mu(v-v')_\mu = 2(v\cdot v' - 1)$. Therefore, in the heavy quark symmetry limit the hadronic current can be written as

$$\langle P'(v')|\bar{h}_{v'}\gamma^\mu(1-\gamma^5)h_v|P(v)\rangle = \xi(v\cdot v')(v+v')^\mu. \quad (2.86)$$

This equation defines $\xi(v\cdot v')$ and gives a relation in the HQS limit between the form factors $f_+(q^2)$ and $f_-(q^2)$ governing the pseudoscalar to pseudoscalar transition, see Eq. 1.5.

The next step is to relate the form factors for the pseudoscalar to vector transition to the form factor $\xi(v\cdot v')$. First, a way of relating the pseudoscalar to the vector is needed. Denote the spin of the heavy quark in the meson by either $\Uparrow$ or $\Downarrow$ depending on whether the spin is up or down. Similarly, denote the spin of the light quark by $\uparrow$ or $\downarrow$. Using these basis states the standard set of basis vectors for the scalar and the vector are written down when coupling a pair of spin 1/2 particles

$$P = \frac{1}{\sqrt{2}}(\Uparrow\downarrow + \Downarrow\uparrow), \quad (2.87)$$

$$
\begin{aligned}
V(\varepsilon_1) &= \Uparrow\uparrow, & (2.88)\\
V(\varepsilon_2) &= \Downarrow\downarrow, & (2.89)\\
V(\varepsilon_3) &= \frac{1}{\sqrt{2}}(\Uparrow\downarrow - \Downarrow\uparrow). & (2.90)
\end{aligned}
$$

Let S_Q^3 be the operator acting on the heavy quark spin such that $S_Q^3 \Uparrow = \frac{1}{2} \Uparrow$ and $S_Q^3 \Downarrow = -\frac{1}{2} \Downarrow$. Applying this operator to P gives the interesting result

$$
S_Q^3 P = \frac{1}{2} V(\varepsilon_3). \tag{2.91}
$$

This shows that applying the heavy-quark operator S_Q^3 gives a relation between the pseudoscalar state and the vector state with longitudinal polarization. It follows that a general matrix element can be written as

$$
\begin{aligned}
\langle V'(v', \varepsilon_3)|\bar{h}'_{v'}\Gamma h_v|P(v)\rangle &= \langle P'(v')|2S_{Q'}^3 \bar{h}'_{v'}\Gamma h_v|P(v)\rangle \\
&= \langle P'(v')|\bar{h}'_{v'}(2S^3\Gamma)h_v|P(v)\rangle \qquad (2.92)
\end{aligned}
$$

where $S^3 = \frac{1}{2}\gamma^5 \not{v}' \not{\varepsilon}^3$ is a matrix representation of the spin operator $S_{Q'}^3$ and Γ is an arbitrary combination of Dirac gamma matrices. The calculation is easiest to perform in the rest frame of the final state meson, where $v'^\mu = (1,0,0,0)$ and $\varepsilon_3^\mu = (0,0,0,1)$. In this frame $S^3 = \frac{1}{2}\gamma^5\gamma^0\gamma^3$. In weak decays Γ will be on the form $V - A$ or $\gamma^\mu(1-\gamma^5)$ and it is simple to evaluate $2S^3\Gamma$ is in the four cases

$$
\begin{aligned}
2S^3\gamma^0(1-\gamma^5) &= \gamma^3(\gamma^5 - 1), & (2.93)\\
2S^3\gamma^1(1-\gamma^5) &= -i\gamma^2(\gamma^5 - 1), & (2.94)\\
2S^3\gamma^2(1-\gamma^5) &= i\gamma^1(\gamma^5 - 1), & (2.95)\\
2S^3\gamma^3(1-\gamma^5) &= \gamma^0(\gamma^5 - 1). & (2.96)
\end{aligned}
$$

These relations are easily verified by explicit calculations using the gamma matrix commutation relations and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

In terms of four-velocities the general matrix element for a pseudoscalar to vector meson can be written as

$$
\begin{aligned}
&\langle V'(v', \varepsilon)|\bar{h}'_{v'}\gamma^\mu(1-\gamma_5)h_v|P(v)\rangle = \\
&\quad i\epsilon^{\mu\nu\alpha\beta}\varepsilon^*_\nu v'_\alpha v_\beta \xi^g(v\cdot v') + \varepsilon^{*\mu}\xi^f(v\cdot v') + \\
&\quad \varepsilon^{*\alpha}v_\alpha v^\mu \xi^a(v\cdot v') + \varepsilon^{*\alpha}v_\alpha v'^\mu \xi^b(v\cdot v') \qquad (2.97)
\end{aligned}
$$

where ξ^g, ξ^f, ξ^a, and ξ^b are invariant form factors. This expression is now evaluated in V''s rest frame with $\varepsilon^\mu = \varepsilon_3^\mu = (0,0,0,1)$. This yields the following four equations

$$\begin{aligned}
\xi(v\cdot v') &= v^0\xi^a(v\cdot v')+\xi^b(v\cdot v'), && (2.98)\\
v^2\xi(v\cdot v') &= v^2\xi^g(v\cdot v')+iv^2v^1\xi^a(v\cdot v'), && (2.99)\\
v^1\xi(v\cdot v') &= v^1\xi^g(v\cdot v')-iv^3v^2\xi^a(v\cdot v'), && (2.100)\\
\xi(v\cdot v')(v^0+1) &= -\xi^f(v\cdot v')+(v^3)^2\xi^a(v\cdot v'). && (2.101)
\end{aligned}$$

From these relations it is easy to solve for the form factors.

$$\begin{aligned}
\xi^g(v\cdot v') &= \xi(v\cdot v'), && (2.102)\\
\xi^f(v\cdot v') &= -\xi(v\cdot v')(v\cdot v'+1), && (2.103)\\
\xi^a(v\cdot v') &= 0, && (2.104)\\
\xi^b(v\cdot v') &= \xi(v\cdot v'). && (2.105)
\end{aligned}$$

Putting this together the weak hadronic current can be written in the form

$$\begin{aligned}
&\langle V'(v',\varepsilon)|\bar{h}'_{v'}\gamma^\mu(1-\gamma_5)h_v|P(v)\rangle = \\
&\quad i\epsilon^{\mu\nu\alpha\beta}\varepsilon^*_\nu v'_\alpha v_\beta\xi(v\cdot v')-\varepsilon^{*\mu}(v\cdot v'+1)\xi(v\cdot v')+\\
&\quad \varepsilon^{*\alpha}v_\alpha v'^\mu\xi(v\cdot v'). \qquad (2.106)
\end{aligned}$$

From this, the relations between the invariant form factors are derived

$$\xi(v\cdot v') = R^*V(q^2) = R^*A_2(q^2) = R^*[1-\frac{q^2}{(M+M')^2}]^{-1}A_1(q^2), \quad (2.107)$$

where $R^* = 2\sqrt{MM'}/(M+M')$.

These relations between the form factors are the most important assumptions used in the form factor measurement presented in this dissertation. Note that HQS does not predict that the q^2 dependence is the same for all form factors.

The other extremely important prediction of HQS is the absolute normalization of $\xi(v\cdot v')$ at the zero recoil configuration, $\xi(v\cdot v'=1)=1$. However, this result is not important for the measurement of the form factors since the fit is only sensitive to the shape and not the absolute normalization. It is of course crucial in the determination of $|V_{cb}|$ using the q^2 distribution.

2.6.2 Heavy quark effective theory

The previous section presented some important results of heavy quark symmetry. The interest of this work is to study the process $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ in which the heavy quark transition is $b \to c$. Of course neither bottom or charm quarks are infinitely heavy. This leads to corrections to the heavy quark symmetry limit. The framework in which these corrections are carried out is called heavy quark effective theory, HQET. This is a systematic expansion of the QCD Lagrangian in terms of the parameter E/m_Q. m_Q is the mass of the heavy quark and E is a relevant energy scale, e.g., Λ_{QCD}. Neubert [24] has carried out these correction to second order in $1/m_Q$. The calculation also includes corrections to order α_s.

Neubert defines the two form factor ratios $R_1(w)$ and $R_2(w)$

$$R_1(w) \equiv [1 - \frac{q^2}{(M_B + M_{D^*})^2}]\frac{V(q^2)}{A_1(q^2)}, \tag{2.108}$$

$$R_2(w) \equiv [1 - \frac{q^2}{(M_B + M_{D^*})^2}]\frac{A_2(q^2)}{A_1(q^2)}. \tag{2.109}$$

These two ratios are both unity in the heavy quark symmetry limit. However, the corrections, as calculated, by Neubert [24] are not negligible

$$R_1(w) = 1.35 - 0.22(w-1) + 0.09(w-1)^2, \tag{2.110}$$

$$R_2(w) = 0.79 + 0.15(w-1) - 0.04(w-1)^2. \tag{2.111}$$

Note that the w dependence of the ratios is very small, the range of accessible w is from 1.0 to 1.5. These corrections have also been evaluated by Close and Wambach [25] who found

$$R_1(w) = 1.15 - 0.07(w-1), \tag{2.112}$$

$$R_2(w) = 0.91 + 0.04(w-1). \tag{2.113}$$

The two calculations agrees on the general trend that R_1 is greater than 1 and that R_2 is less than 1. They also agree in the prediction of a mild w dependence.

The other interesting thing about the $1/m_Q$ corrections is the effect they have on the absolute normalization of $\xi(w)$. According to what is known as Luke's theorem [26] the $1/m_Q$ corrections vanishes at $w = 1$, and the largest corrections come from the second to leading order contributions.

2.7 Functional form in the form factor fit

After discussing the invariant form factors, the helicity amplitudes, the form of the decay rate, and heavy quark symmetry, the functional form that is used in the fit for the form factors can now be stated and motivated. As mentioned earlier, the fit is a three parameter fit to two form factor ratios and a slope parameter that parametrizes the q^2 dependence of the form factors.

The slope parameter, $\rho^2_{A_1}$, is the slope of the Isgur-Wise function. However, the Isgur-Wise function is not a real form factor since it only exists in the limit of infinitely heavy quark masses. Instead of using the Isgur-Wise function directly, the form factor $h_{A_1}(w)$ is introduced. This form factor is related to the invariant form factor A_1 by

$$h_{A_1}(w) = R^*[1 - \frac{q^2}{(M_B + M_{D^*})^2}]^{-1} A_1(q^2). \tag{2.114}$$

As the charm and bottom quark masses go to infinity the form factor $h_{A_1}(w) \to \xi(w)$. This form factor is parametrized by

$$h_{A_1}(w) = h_{A_1}(1)[1 - \rho^2_{A_1}(w-1)], \tag{2.115}$$

where $\rho^2_{A_1}$ is refered to as the form factor slope or simply the slope. $h_{A_1}(1)$ is the overall normalization. This measurement is not sensitive to the normalization and therefore $h_{A_1}(1)$ is not relevant for this measurement. However, in a measurement of $|V_{cb}|$, $h_{A_1}(1)$ has to be given by theoretical predictions, since such a measurement is sensitive to the product $|V_{cb}h_{A_1}(1)|$.

Given $h_{A_1}(w)$ the other two form factors can be expressed as

$$V(q^2) = R_1(w){R^*}^{-1}h_{A_1}(w), \tag{2.116}$$
$$A_2(q^2) = R_2(w){R^*}^{-1}h_{A_1}(w). \tag{2.117}$$

These two equations defines the two form factor ratios that are measured in this analysis. These ratios are taken to be constant, so the extracted parameters from the fit are R_1, R_2, and $\rho^2_{A_1}$.

A set of alternative fits are also performed to test the sensitivity to the assumptions made above. First alternative functional forms of $h_{A_1}(w)$ are used. Table 2.1 summarizes the different forms used for $h_{A_1}(w)$. To test the sensitivity to the assumption that R_1 and R_2 are constants fits are performed assuming the w dependence on R_1 and R_2 given by Eq. 2.111 and Eq. 2.113.

$h_{A_1}(w)/h_{A_1}(1)$	Expansion
$1-\rho^2_{A_1}(w-1)$	$1-\rho^2_{A_1}(w-1)$
$\exp[-\rho^2_{A_1}(w-1)]$	$1-\rho^2_{A_1}(w-1)+\rho^4_{A_1}\dfrac{(w-1)^2}{2}$
$\frac{2}{w+1}\exp[(1-2\rho^2_{A_1})\frac{w-1}{w+1}]$	$1-\rho^2_{A_1}(w-1)+(\rho^4_{A_1}+\rho^2_{A_1}-\frac{1}{4})\dfrac{(w-1)^2}{2}$
$(\frac{2}{w+1})^{2\rho^2_{A_1}}$	$1-\rho^2_{A_1}(w-1)+(\rho^2_{A_1}+\frac{1}{2}\rho^2_{A_1}(2\rho^2_{A_1}-1))\dfrac{(w-1)^2}{2}$

Table 2.1: The alternative forms used for $h_{A_1}(w)$ and their Taylor expansion around $(w-1)$ to second order. Note that all forms has the same linear term, i.e., that $\rho^2_{A_1}$ is the slope at $w=1$.

Chapter 3

The CLEO II experiment

The data used in the analysis of the $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ form factors were collected by the CLEO II experiment. The data set used is an excellent source of B-mesons for this study; the B-mesons are produced in decays of the $\Upsilon(4S)$ which gives them only a small momentum, about 325 MeV/c. This is used as a constraint in the analysis to enhance the cleanliness of the signal.

Figure 3.1 shows the cross-section for $e^+e^- \to$ hadrons as a function of the center of mass energy, $\sqrt{s}$, around the mass of the $\Upsilon(4S)$. The $\Upsilon(4S)$ is the lightest $b\bar{b}$ resonance that is kinematically allowed to decay to a pair of B-mesons. The cross-section for $e^+e^- \to \Upsilon(4S)$ is ≈ 1.07 nb at the peak of the $\Upsilon(4S)$, $\sqrt{s} = 10.58$ GeV. The cross section for continuum $q\bar{q}$-production is approximately 3 nb. Data is also collected at an energy 50 MeV below the $\Upsilon(4S)$. These data are used to model the continuum background at the $\Upsilon(4S)$.

The CLEO II detector, shown in side view in Figure 3.2, is a typical solenoidal collider detector. Closest to the interaction point, IP, are three charged particle detectors. These position sensitive detectors forms the tracking system in the CLEO II detector. The tracking system is operated in a 1.5 T solenoidal magnetic field to allow for precise measurements of the charged particle's momentum. The tracking system is described below.

Outside the tracking system is the time of flight (ToF) consisting of 64 barrel counters and 28 counters on each endcap. This device measures the time, from a beam collision, untill the particle hits the ToF detector. This measurement of the time or equivalent the speed of the particle combined with the momentum measurement from the tracking devices provide a mean of particle identification. The information from the ToF system is not used in the form factor analysis described in this dissertation.

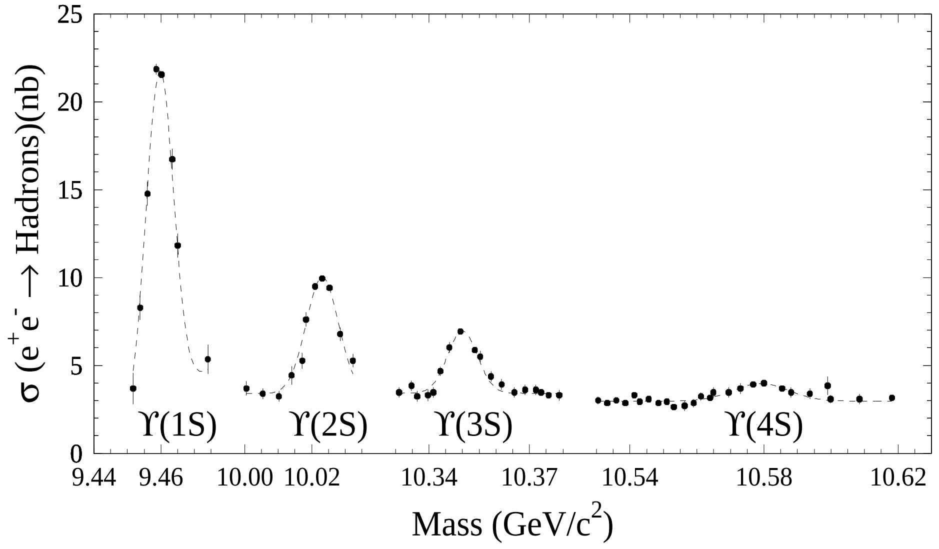

Figure 3.1: The hadronic cross section for $e^+ + e^-$ at and around the $\Upsilon(4S)$.

Further outside is the electromagnetic calorimeter, EMC. This detector consists of 7800 CsI crystals. They measure energies deposited by electromagnetically interacting particles, in particular electrons and photons. Besides measuring energies of photons the energy measurement of electrons provide an excellent electron identification since $E/P = 1$.

Outside the EMC is the muon system. Muons are identified by their ability to penetrate material without being absorbed. The absorption material in the CLEO II detector is steel, it also serves as the return yoke for the magnetic field. A measure of how likely a particle is to be a muon is obtained by measuring how many nuclear interaction lengths it passed through, and still is detected.

3.1 The tracking system

As shown in Figure 3.2 the tracking system consists of 3 tracking devices. The precession tracking layers, PTL, is closest to the IP and has the best resolution of the tracking devices. It provides the vertex information, i.e., the information about the particles production positions. The PTL is a six layered straw-tube wire chamber with 64 wires per layer. It has a resolution of ≈ 50 μm on the measured drift distance. The layers are staggered to help

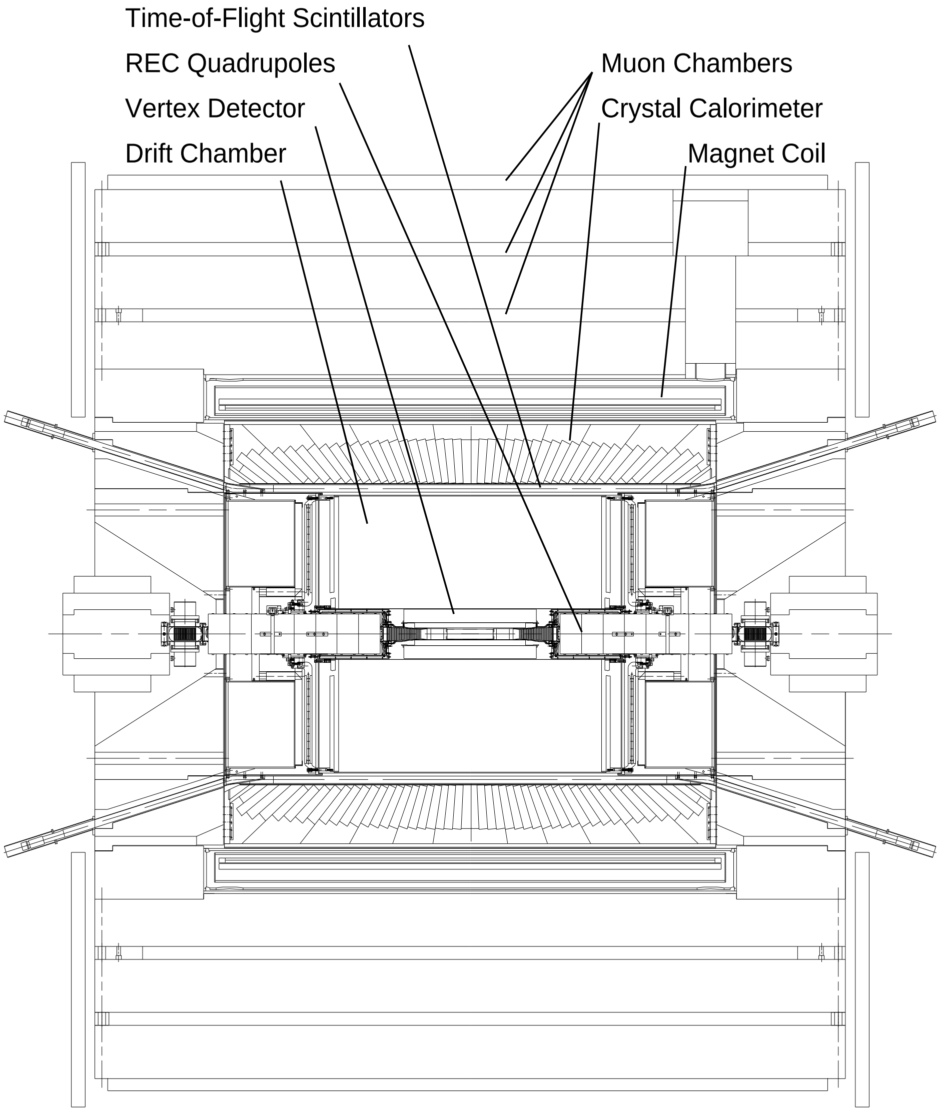

Figure 3.2: The CLEO II detector in a side view.

solving the sign ambiguity.

The next tracking detector is the vertex detector[1]. The VD is a 10 layered drift chamber with 64 sense wires in the innermost layer and 96 in the outermost.

The DR is the main drift chamber. It has a total of 51 layers. There are 41 axial layers and 10 stereo layers. The stereo layers provide information about the z-position of a track. The inner radius of the DR is 25 cm and its outer radius is 90 cm. The drift chamber provides information on the curvature of the track. It also provides information on the z-position of the track. Besides the stereo wires information on the z-position of the track there are "cathode" measurements of the z-position. The cathodes are pads on the inner and outer walls of the VD and the DR sensing the z-position of a track as it enters and exits the detectors.

The information from these detectors are used in the tracking software to find the tracks, pattern recognition, and to fit for the particles momentum and production vertex, track fitting. For more information on track fitting see Part II of this dissertation.

The drift chamber also provides particle identification information through measurements of the specific ionization. This information is used in the analysis to discriminate kaons from pions, see Section 4.2.2.1.

3.2 The electromagnetic calorimeter

In this analysis the EMC is used for two different purposes. First, it is used for electron identification; an electron deposits all energy in the calorimeter and by matching tracks and showers with $E/P = 1$ electrons are identified. In the analysis electrons are identified as having $R_e > 3.0$ where

$$R_e = \sum \ln \frac{P_e}{P_{\not{e}}}. \tag{3.1}$$

P_e is the probability for the particle identification to identify the particle as an electron while $P_{\not{e}}$ is the probability that it is not an electron, i.e., it is either a pion, moun, kaon, or proton. The sum runs over the different ways of identifying particles, dE/dx, ToF, E/P, etc. However, for the momentum range $P_e > 1.0$ GeV/c, the information from E/P is dominant.

[1]The name is an anachronism from the time when the vertex detector was the detector closest to the IP and provided the vertexing.

The second use of the EMC is π^0 reconstruction. Due to the excellent energy resolution, $\sigma_E = 5\%$ at 100 MeV and $\sigma_E = 1.5\%$ at 5 GeV, the CLEO detector can reconstruct π^0's with a mass resolution of about 5 MeV/c^2.

3.3 Muon system

The return yoke of the magnet serves as the muon identification system. The yoke is instrumented with proportional counters at depths of 36, 72, and 108 cm. The equivalent thickness in nuclear interaction length varies from 7.2 to 10 depending on the direction of the track. To obtain a high effeciency the proportional counters in the gaps of the iron are three layered.

In this analysis muons are identified as tracks that passes through at least 5 nuclear interaction lengths. This gives an effective cut-off of P_μ less than 1.2–1.3 GeV/c for finding muons. In the analysis the momentum cut on the muons are raised to 1.4 GeV/c. The requirement of passing through 5 nuclear interaction lengths gives a fake probability, mainly from pions and kaons, of about 2.5% per hadron.

3.4 The data set used

The data used in this analysis were collected in the period from November 1990 to February 1994. The data are devided into 9 data sets, labeled 4S2 through 4SA. The luminosities and other information about these data sets are summarized in Table 3.1. There are no major changes between these data sets. The most significant change is the use of DME gas in the PT starting in 4S4. Otherwise only the effects of aging are of any significance.

Data set	Data taken	on res lum. (pb^{-1})	off res lum. (pb^{-1})
4S2	Nov 11-90 – Jun 04-91	465	197
4S3	Sep 18-91 – Feb 17-92	436	208
4S4	Apr 10-92 – May 26-92	214	101
4S5	Jul 09-92 – Oct 15-92	216	105
4S6	Nov 03-92 – Jan 19-93	233	85
4S7	Mar 16-93 – Jul 06-93	285	177
4S8	Aug 01-93 – Sep 27-93	189	94
4S9	Nov 22-93 – Jan 10-94	230	116
4SA	Jan 20-94 – Feb 28-94	138	54
Total		2404	1137

Table 3.1: Luminosities of the data sets.

Chapter 4

Event selection

The measurement of the form factors is greatly simplified by having a clean signal with a small contamination from backgrounds. The selection criteria applied to obtain the event sample used in the form factor fits are discussed below. Before discussing the event selection, a brief introduction to the sources of background events is given. This will help the discussion of the event selection criteria.

The event selection criteria in this analysis are quite stringent, since the background distributions must be understood in the four dimensions used in the fit. By rejecting most background events the sensitivity to uncertainties in the background shape is reduced.

4.1 Discussion about backgrounds

The backgrounds in this analysis are simply broken down into a few well defined components. Most background events both in the $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^+\pi^0$ channels, are fake D^*s, which are combinatorial background under the D^* peak. This background is estimated using the sideband of the D^*–D mass difference (δm) distribution. The second largest background is feed through from $B \to D^{**}\ell\nu$ where $D^{**} \to D^{*+}\pi$. The estimate of this background is based on a combination of measurements of branching fractions for $B \to D^{**}\ell\nu$ by ALEPH [10] and predictions of the ISGW2 [23] model. Another background is due to events in which a correctly reconstructed D^* and a real lepton candidate are selected, but the two particles come from different Bs. For this to happen, the event either underwent $B^0\bar{B}^0$-mixing or the lepton is secondary, since otherwise the charge correlation between the D^* and the lepton is wrong. This background is called uncorrelated since the direction of the D^* and the lepton are uncorrelated (the Bs are produced essentially at rest). Since the D^* and the lepton are uncorrelated, this background can be estimated using a procedure in which the direction of the lepton momentum is reversed. Another background is due to fake leptons. This background mainly contributes to the muon channel, and is estimated using measured fake rates. The last background of any significance is the contribution from nonresonant $c\bar{c}$ production at the $\Upsilon(4S)$. This background is estimated using data taken below the energy of the $\Upsilon(4S)$ resonance.

4.2 Event selection

After identifying the main backgrounds in this analysis the event selection criteria are discussed. First the selection criteria for the leptons are discussed and then the reconstruction of the D^0 and the D^{*+} is considered. After all particles are reconstructed event selection criteria based on kinematic consistency are considered.

4.2.1 Lepton selection

Hadronic events [27] are selected with a track identified as an electron in the momentum range 1.0 GeV/c to 2.45 GeV/c or a muon in the range 1.4 GeV/c to 2.45 GeV/c. There is a series of constraints that decide the momentum range for the leptons. The high energy cutoff at 2.45 GeV/c is the kinematic endpoint. For muons, there is an effective cut-off at 1.4 GeV/c from the muon identification requirement that the charged tracks penetrate at least 5 nuclear interaction length. The electron identification, essentially E/P, is effective down to about 0.5 GeV/c. But here the lower cut is determined by another constraint: at low lepton energies the contribution from $B \to D^{**}\ell\nu$ is significant while the signal is vanishing. These things together motivate the choice of 1.0 GeV/c as the lower energy cut for the electrons. A track is identified as an electron when the likelihood ratio R_e (see Section 3.2) is greater than 3.0. Both the electrons and muons are required to be in the good barrel region, i.e., $|\cos\theta| < 0.71$, where θ is the angle between the beam axis and the lepton momentum.

4.2.2 D^0 reconstruction

Originally this analysis was performed only in the decay channel $D^0 \to K^-\pi^+$, but the results were limited by the statistical error. The decay mode $D^0 \to K^-\pi^+\pi^0$ was therefore investigated as an additional mode to increase the statistical power of the form factor measurement. After using a series of event selection criteria, described below, a signal in the $D^0 \to K^-\pi^+\pi^0$ channel is obtained with only slightly more background than in the original $D^0 \to K^-\pi^+$ channel.

There is one further difference between the two modes. In the $D^0 \to K^-\pi^+\pi^0$ mode only one combination per event is selected. Multiple entries per event in the $D^0 \to K^-\pi^+\pi^0$ channel are almost entirely due to photon combinatorics in forming the π^0 candidates; quite often a hard photon makes several combinations with different softer photons. Only $\approx 5\%$ of the events in the $D^0 \to K^-\pi^+\pi^0$ channel have more than one combination and in fact quite often more than two. If multiple combinations pass all event-selection criteria, the combination with $m_{\gamma\gamma}$ closest to the nominal π^0-mass is selected. If the combinations have the same mass, i.e., the same photon candidates, the combination with the largest Dalitz probability (see below) is selected.

4.2.2.1 The $D^0 \to K^-\pi^+$ channel

The D^0 candidates are formed from pairs of tracks with opposite charges and the track with the same charge as the lepton is assigned the kaon mass. No particle identification is used for these tracks.

The D^0 candidates are required to have a momentum less than 2.45 GeV/c, which is close to the kinematic threshold for D^0s coming from B decays. This requirement helps suppress background from nonresonant charm production, which has a harder momentum spectrum. The D^0 candidates are required to be within 16 MeV/c^2 of the nominal D^0 mass, which is approximately a 2σ cut. Figure 6.4 shows the $M_{K\pi}$ distribution after all event selection criteria are applied except the criterion on $M_{K\pi}$.

4.2.2.2 The $D^0 \to K^-\pi^+\pi^0$ channel

π^0 candidates are formed from showers in the calorimeter that are not matched to a charged track. The π^0s used are required to have individual showers, corresponding to photons, with energies of at least 50 MeV. At least one of the two photons is required to be in the "good barrel" region, $|\cos\theta_\gamma| < 0.71$, where θ_γ is the angle of the photon with respect to the beam axis. To obtain optimal momentum resolution of the π^0, the two photons are kinematically fit with the π^0 mass as a constraint. π^0 candidates are accepted if they are within 2σ of the nominal π^0 mass and have an energy of at least 350 MeV.

The π^0 candidates are combined with two oppositely charged tracks, the kaon and pion candidates, to form the D^0 candidate. The track with the same charge as the lepton is assigned the kaon mass. Since the kaon momentum spectrum is soft, there is useful information in the particle identification; dE/dx is used and it is required that $\sigma_K > -2.0$[1]. This selection criteria rejects pions from being accepted as kaon candidates; this is the biggest misidentification problem, as seen in Figure 4.1.

To reduce the background further, a selection criterion based on the Dalitz probability[2], $\mathcal{D}$, of the decay $D^0 \to K^-\pi^+\pi^0$ is used. The resonant substructure of this decay has been measured by the E691 collaboration [28]. The Dalitz plot, shown in Figure 4.2, is dominated by the $K^*(892)$ and $\rho(770)$ resonances. This strong substructure makes a cut on the Dalitz probability efficient. D^0 candidates with $\mathcal{D} > 12.8\%$ are selected, which means that the events with a Dalitz probability greater than 12.8% of the maximum are accepted. This selection criterion is efficient in reducing the background from fake Ds, as seen in Figure 4.3. The efficiency for this criterion is 62% for the signal and 29% for the background (modeled by the D^0 sideband).

D^0 candidates with a mass within 25 MeV/c^2 (2σ) of the nominal D^0 mass are accepted. This cut is looser than in the $D^0 \to K^-\pi^+$ analysis, where the mass resolution is 8 MeV/c^2, as compared to 13 MeV/c^2 in the $D^0 \to K^-\pi^+\pi^0$ channel.

4.2.3 D^* reconstruction

The D^0 candidates are combined with another pion, the slow pion, to form a D^* candidate. This pion is required to have a momentum of at least 65 MeV/c and to be in the good barrel region of the detector, $|\cos\theta_{\pi_s}| < 0.71$, where θ_{π_s} is the angle of the slow pion momentum with respect to the beam line. The reconstructed D^*–D mass difference, $\delta m = M_{D^*} - M_{D^0}$ is then formed. The signal region is defined to be the region within 2 MeV/c^2 of the nominal mass difference of 145.5 MeV/c^2. To study the background from

[1] σ_K is the number of standard deviations away from the most likely value the measured dE/dx is for a kaon.

[2] $\mathcal{D}$ is the amplitude squared for the decay $D^0 \to K^-\pi^+\pi^0$.

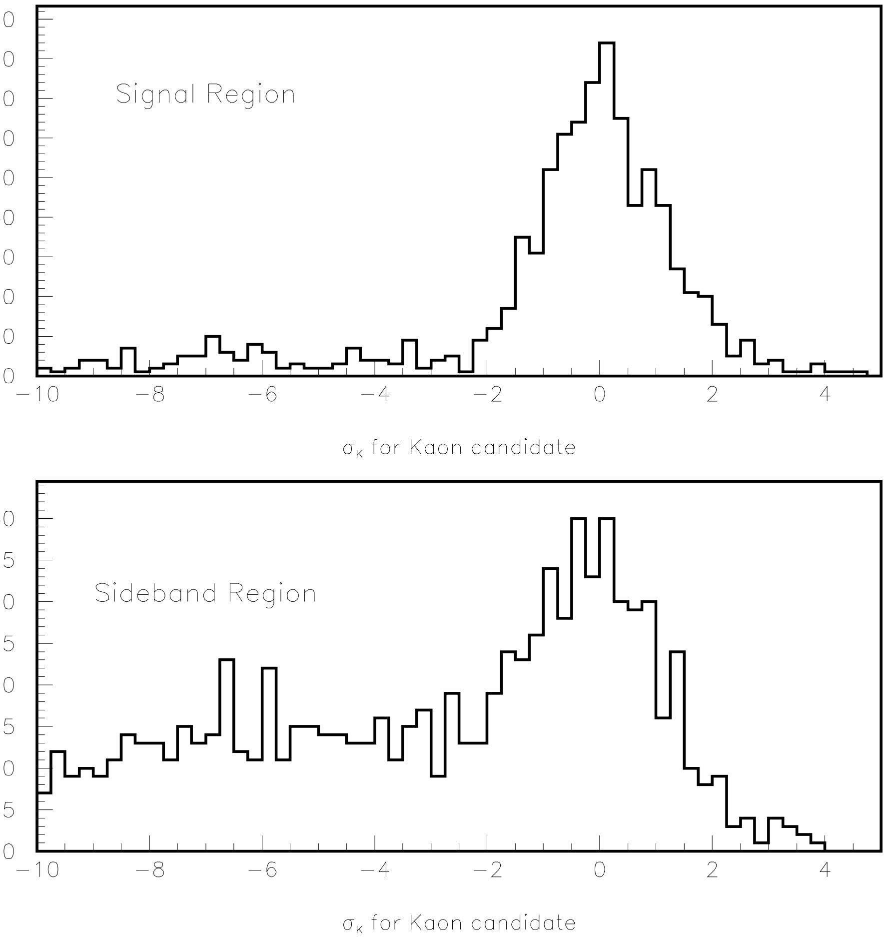

Figure 4.1: The σ_K distribution in data for the signal region ($|M_{K\pi\pi^0} - M_{D^0}| < 25$ MeV/c^2) and the side band ($|M_{K\pi\pi^0} - 1737.5| < 62.5$ MeV/c^2) in D^0 mass. The plots are made without the cut applied on the Dalitz probability.

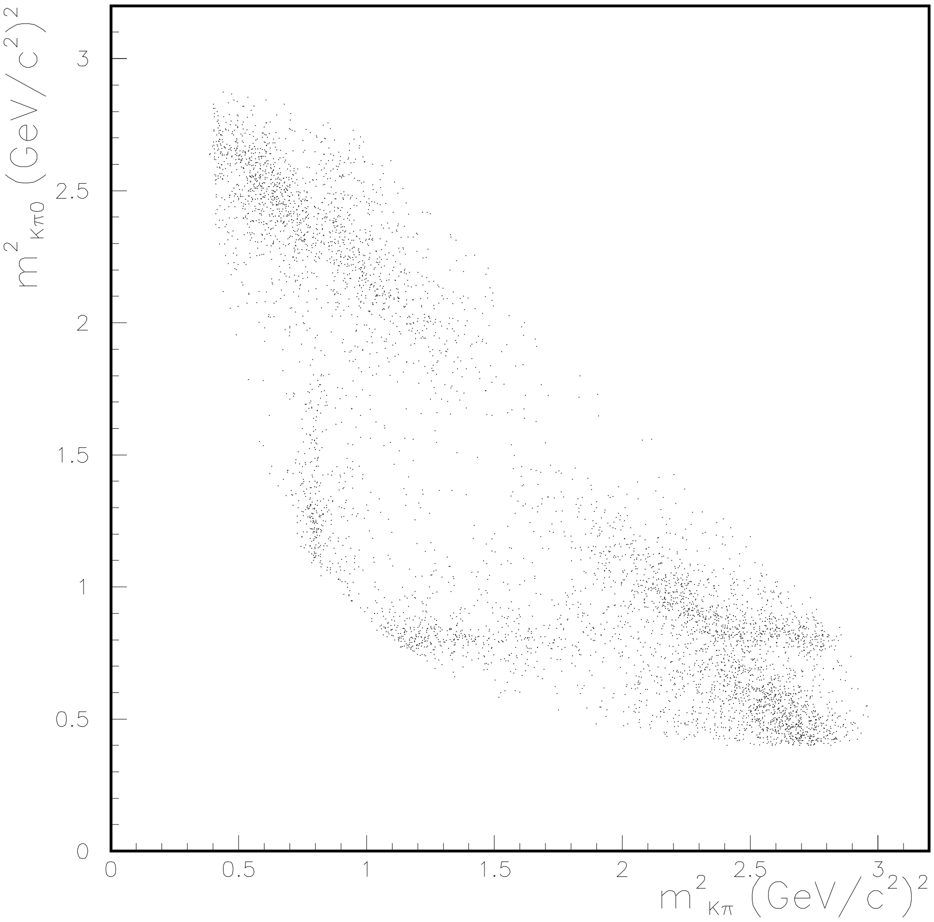

Figure 4.2: The Dalitz plot for the decay $D^0 \to K^-\pi^+\pi^0$. The horizontal axis is $M^2_{K^-\pi^+}$ while on the vertical axis $M^2_{K^-\pi^0}$ is plotted. The horizontal and vertical bands in the Dalitz plot are due to the charged and neutral $K^*(890)$ resonances. The diagonal band is the $\rho(770)$ resonance. The plot is for signal Monte Carlo.

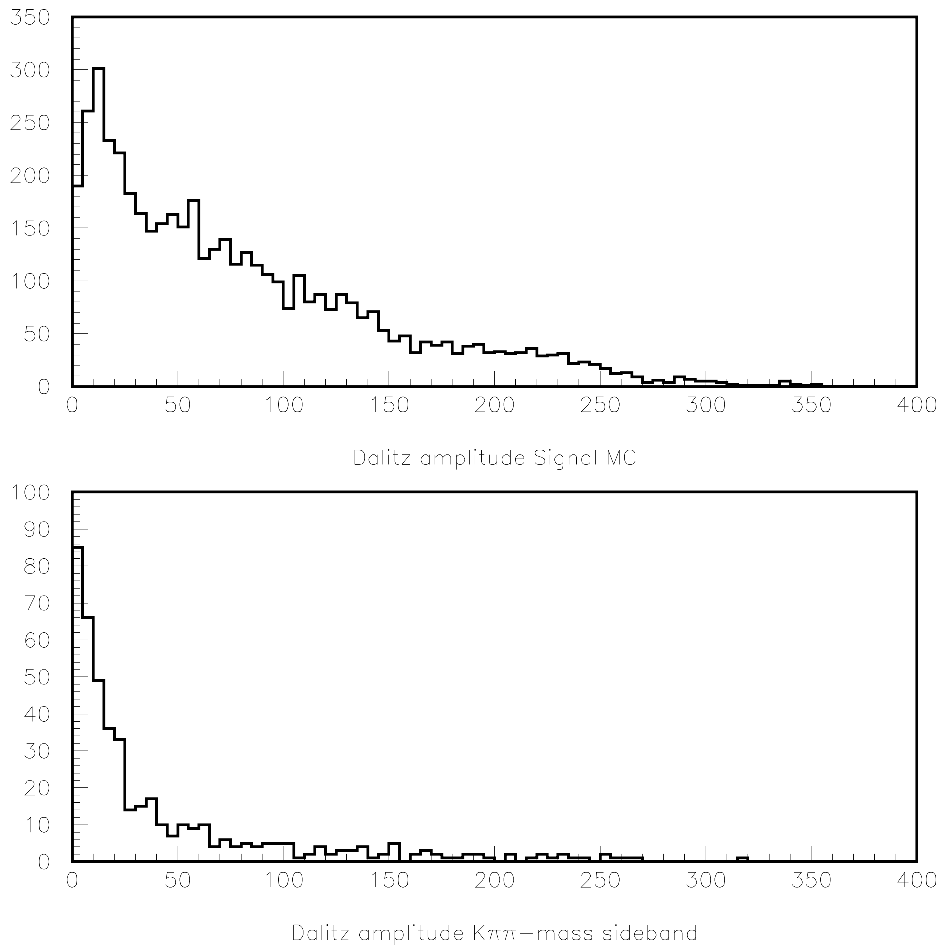

Figure 4.3: The Dalitz probability for the signal, from Monte Carlo, and from the D^0 sideband, from data. The D^0 sideband is taken as a model for the background, i.e., wrong $K\pi\pi^0$ combinations. The background is more strongly peaked at zero than the signal. Events with $\mathcal{D} > 12.8\%$ or at 45 in the arbitrary scale in the figure are selected. This criterion is 29% efficient on the D^0 sideband, while it retains 62% of the signal.

fake D^*s the δm sideband region from 150 MeV/c^2 to 170 MeV/c^2 is used. In Figure 4.5 and 4.6 the δm distributions are shown in the $D^0 \to K^-\pi^+$ and the $D^0 \to K^-\pi^+\pi^0$ mode after all cuts except the cut on δm. The $D^0 \to K^-\pi^+\pi^0$ mode is almost as clean as the $D^0 \to K^-\pi^+$ mode.

4.2.4 Kinematic consistency

Beyond the selection criteria discussed above, event selection criteria based on the kinematic properties of the semileptonic decay and the slow moving B-meson are applied. There is also an event selection criterion based on the global event shape.

4.2.4.1 Global event shape, R_2

The global event shape variable R_2 is defined to be the ratio of the Fox-Wolfram moments [29] H_2 and H_0. R_2 is required to be less than 0.3, which rejects jet-like events. This criteria rejects continuum (non-$\Upsilon(4S)$) events since these events tend to have a jet-like topology as compared to the more spherical $B\bar{B}$ events.

4.2.4.2 $\cos\theta_{D^*\ell}$

The D^{*+} candidate is required to be in the opposite hemisphere from the lepton, $\cos\theta_{D^*\ell} < 0$. This criterion helps reduce the background from combinations where the lepton and the D^{*+} come from different Bs, i.e., uncorrelated background. The signal is strongly peaked with the momentum of the D^{*+} and the lepton in opposite hemispheres, while the uncorrelated background is uniform in $\cos\theta_{D^*\ell}$, where $\theta_{D^*\ell}$ is the angle, in the laboratory frame, between the D^* and the lepton momentum.

4.2.4.3 $\cos\theta_{BY}$

Using the beam energy constraint, $E_{\text{Beam}} = E_B = E_{D^*\ell} + E_\nu$, the energy and momentum of the inferred neutrino can be calculated. Since $\vec{P}_{\bar{B}^0} = \vec{P}_{D^*\ell} + \vec{P}_\nu$ the three momenta should form a closed triangle. From this, the cosine of the angle, θ_{BY}, between the B and the $D^*\ell$ momentum can be calculated

$$\cos\theta_{BY} = \frac{|\vec{P}_B|^2 + |\vec{P}_{D^*\ell}|^2 - |\vec{P}_\nu|^2}{2|P_B||P_{D^*\ell}|}. \tag{4.1}$$

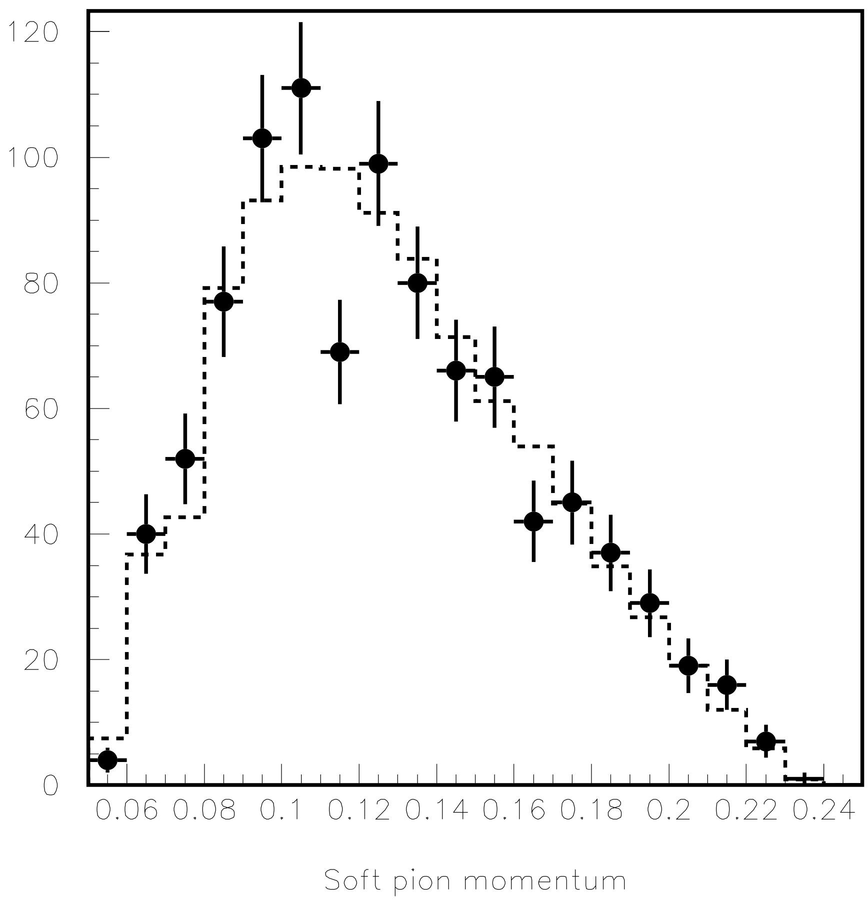

Figure 4.4: The slow pion momentum spectrum in Monte Carlo and data. The solid line is the Monte Carlo and the data are indicated with error bars. This is the spectrum after all cuts in the analysis.

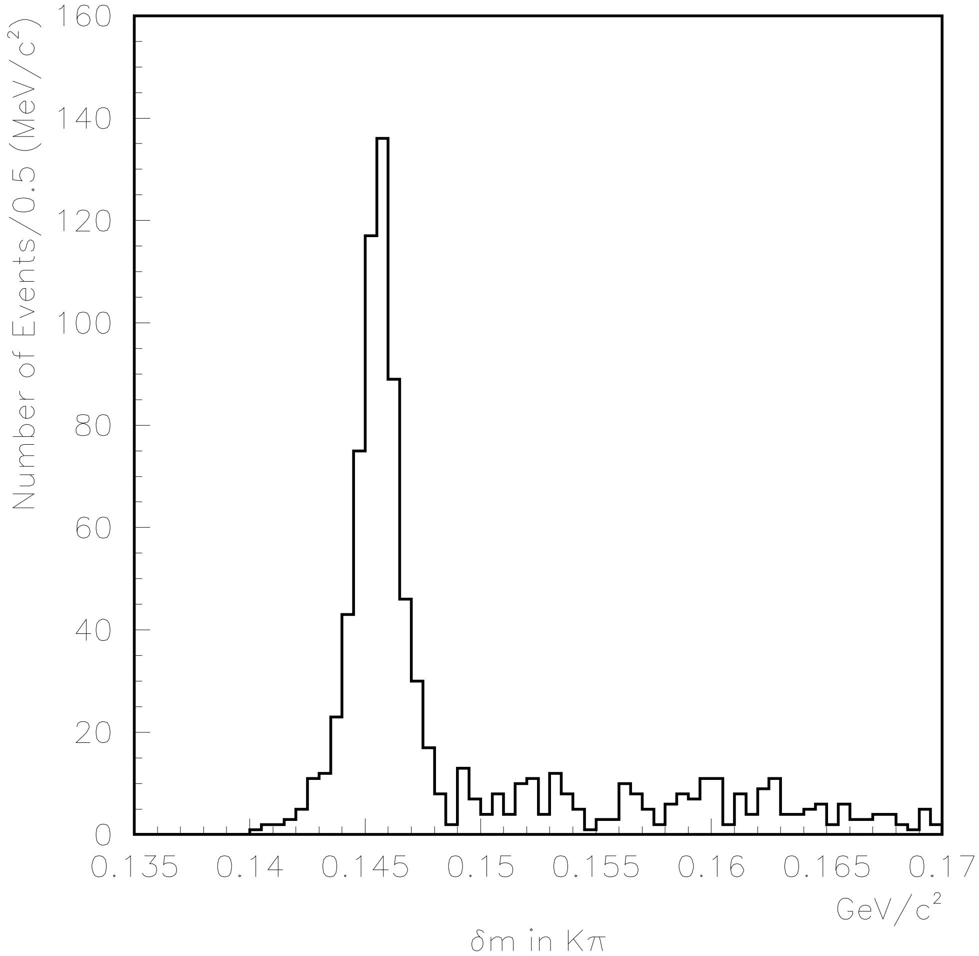

Figure 4.5: The δm mass difference distribution in the $D^0 \to K^-\pi^+$ channel after all cuts except the δm cut.

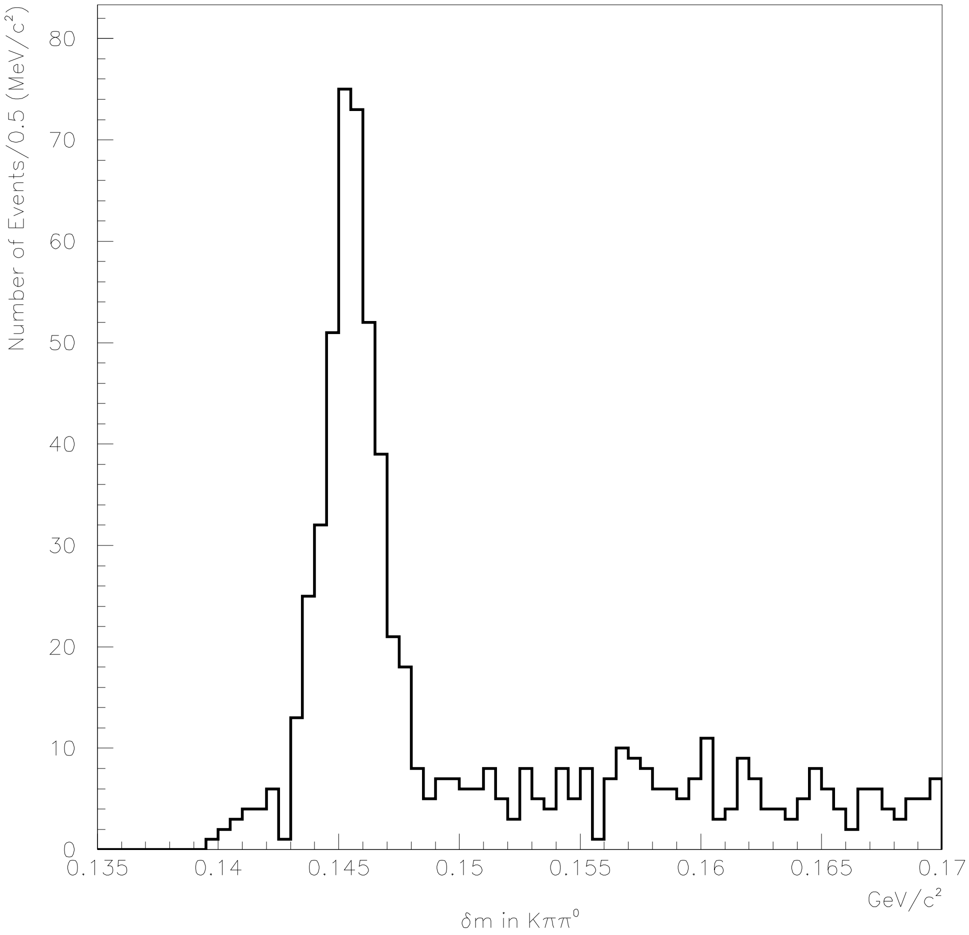

Figure 4.6: The δm mass difference distribution in the $D^0 \rightarrow K^-\pi^+\pi^0$ channel after all cuts except the δm cut.

$\cos\theta_{BY}$ is required to be in the physical range, i.e., $|\cos\theta_{BY}| < 1.0$. $\cos\theta_{BY}$ provides information on the direction of the B: it gives the projection of the B-momentum on the known $D^*\ell$ direction. This will be used below in the calculation of the kinematic variables that describe the decay. In Figs. 4.7 and 4.8 the $\cos\theta_{BY}$ distributions are shown for the electron and muon modes separately. This cut is efficient in rejecting backgrounds in which there are missing particles, e.g., $B \to D^{**}\ell\nu$, $D^{**} \to D^{*+}\pi$ and a pion is missed.

4.2.4.4 The missing mass, $M^2_{\rm miss}$

Another way of imposing the constraint that the event is semileptonic with a missing neutrino is to calculate the missing mass squared

$$M^2_{\rm miss} = (p_B - p_{D^*\ell})^2 = M_B^2 + M^2_{D^*\ell} - 2E_B E_{D^*\ell} + 2\vec{P}_B \cdot \vec{P}_{D^*\ell}. \qquad (4.2)$$

In evaluating $M^2_{\rm miss}$, E_B is taken to be the beam energy and $\vec{P}_B = 0$. If the Bs were produced at rest, $\vec{P}_B = \vec{0}$, the missing mass distribution would be a narrow peak centered at zero. However, the Bs are slowly moving and this leads to a broadening of the $M^2_{\rm miss}$ distribution and also a shift. If the D^* and the lepton are from a $B \to D^{**}\ell\nu$ decay, where an additional pion is missing, the missing mass will be positive. It is required that $M^2_{\rm miss} < 0.0$ for electrons and that $M^2_{\rm miss} < 0.7$ $({\rm GeV/c^2})^2$ for muons. The different cuts for electrons and muons are motivated since different selection criteria are used for the electron and muon momenta. The cuts for electrons and muons are such that the amount of background from $B \to D^{**}\ell\nu$ is of about the same size in both samples.

4.3 Background estimates

The event selection criteria discussed in Section 4.2 strongly suppress the backgrounds. It is particularly important in this measurement of the form factors to have a clean signal since it is hard to understand the background shape in four dimensions.

The background sources in this analysis were briefly discussed in Section 4.1. In the sections below, detailed discussions of the estimates of the backgrounds are given. There are also discussions of how these backgrounds are modeled in the form factor fits. These discussions are continued in Chapter 6 where the systematic uncertainties are described.

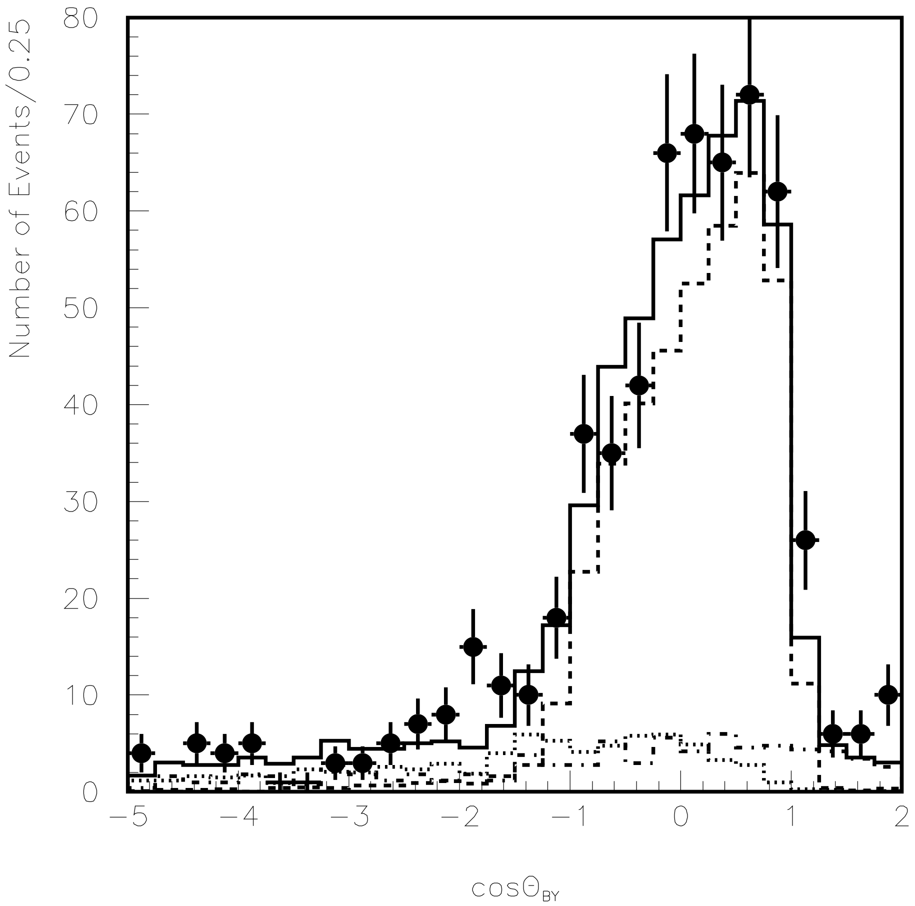

Figure 4.7: The $\cos\theta_{BY}$ distributions after all cuts except the $\cos\theta_{BY}$ cut and missing mass cut for the electrons. The data are indicated with points with error bars. The Monte Carlo signal shape is the dashed line, the background from D^{**} is the dotted line and the normalized δm sideband is the dashed-dotted line. The sum of these contributions is the solid line. It is seen that the solid line, i.e. signal+D^{**}+δm sideband describes the data well. The signal region in this plot is from 0 to 1. The region below 0 is excluded by the requirement that the missing mass should be less than 0 $\mathrm{GeV}^2/\mathrm{c}^4$.

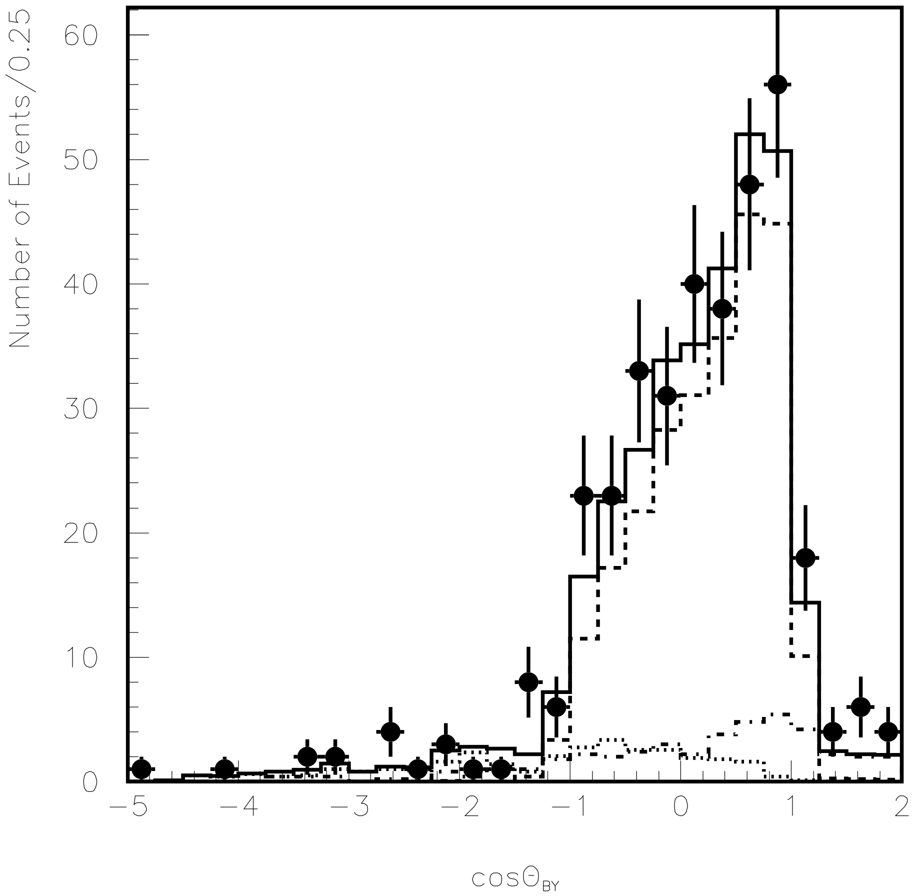

Figure 4.8: The $\cos\theta_{BY}$ distributions after all cuts except the $\cos\theta_{BY}$ cut and missing mass cut for the muons. The data are indicated with points with error bars. The Monte Carlo signal shape is the dashed line, the background from D^{**} is the dotted line, and the normalized δm sideband is the dashed-dotted line. The sum of these contributions is the solid line. The solid line, i.e., signal+D^{**}+δm sideband describes the data well.

$P_\ell(e)$	$1.0 < P_\ell < 2.45$ GeV/c
$P_\ell(\mu)$	$1.4 < P_\ell < 2.45$ GeV/c
P_{D^0}	< 2.45 GeV/c
P_{π_s}	> 65 MeV/c
$\lvert\Delta M_{D^0}\rvert$ $(K\pi)$	<16 MeV/c^2
$\lvert\Delta M_{D^0}\rvert$ $(K\pi\pi^0)$	<25 MeV/c^2
$\lvert\Delta\delta m\rvert$	<2.0 MeV/c^2
$\lvert\cos\theta_{B,D^*\ell}\rvert$	<1.0
R_2	<0.3
$M^2_{\rm miss}$ (e)	< 0.0 GeV^2/c^4
$M^2_{\rm miss}$ (μ)	< 0.7 GeV^2/c^4
$\cos\theta_{D^*\ell}$	<0.0
E_{π^0} $(K\pi\pi^0)$	>350 MeV
$E^{\gamma}_{\rm min}$ $(K\pi\pi^0)$	>50 MeV
$\lvert M_{\gamma\gamma} - M_{\pi^0}\rvert$ $(K\pi\pi^0)$	$< 2\sigma$
σ_K (dE/dx) $(K\pi\pi^0)$	> -2.0
$\mathcal{D}$, $K\pi\pi^0$ probability $(K\pi\pi^0)$	>0.128

Table 4.1: The selection criteria used in the analysis. Cuts specific to a given channel are indicated in parenthesis after the criterion. Besides these cuts, it is required that only one combination passes the selection criteria per event, see Section 4.2.2.

	$K\pi$ e	$K\pi$ μ	$K\pi\pi^0$ e	$K\pi\pi^0$ μ	Total
Raw yield	267 ± 16	292 ± 17	192 ± 14	189 ± 14	940 ± 31
Fake D^*	17 ± 3	21 ± 4	20 ± 4	21 ± 4	79 ± 8
$B \to D^{**}\ell\nu$	12 ± 7	17 ± 8	8 ± 5	11 ± 5	48 ± 25
Continuum	2 ± 2	3 ± 3	0 ± 2	0 ± 2	5 ± 5
Uncorrelated	7 ± 4	3 ± 3	6 ± 3	0 ± 2	16 ± 6
Fake leptons	1 ± 1	6 ± 3	1 ± 1	5 ± 3	13 ± 4
Signal	228 ± 18	242 ± 20	157 ± 15	152 ± 16	779 ± 42

Table 4.2: The yields in the $K\pi$ and $K\pi\pi^0$ modes and the estimated background contamination of the signal. The errors on the amount of the $B \to D^{**}\ell\nu$ background are 100% correlated between the four modes. The errors for the fake D^* background are uncorrelated since they are statistical errors from independent samples.

4.3.1 Fake D^*s

This background is estimated using the δm sideband, which is defined to be the range of δm from 0.15 GeV/c^2 to 0.17 GeV/c^2. The data are used to estimate the amount of this background. As discussed below, Monte Carlo is used to model the shape since there are too few events in the data.

This background is estimated by scaling the δm sideband region to the signal region. There is, however, a small correction for events with high q^2 where the δm distribution is not uniform. The estimated numbers of events in the different modes are summarized in Table 4.2. The background is somewhat higher in the $D^0 \to K^-\pi^+\pi^0$ channel than in the $D^0 \to K^-\pi^+$ channel, but it is still less than 10% of the signal.

This background tends to be peaked at higher q^2, which is easily understood, since high q^2 corresponds to low D^* momentum. In Figures 4.9 and 4.10 the sideband distributions are shown in each of the 4 projections. The few events in the data do not provide a lot of information on the 4-dimensional shape. Therefore, the approach is to generate a high statistics Monte Carlo sample that is used to get the shape of this background for the form factor fits. In Figures 4.9 and 4.10 the dashed line is the high statistics Monte Carlo sample. TRKSIM [30] is used to generate this sample since it requires ≈ 100 million events to collect enough statistics after the event selection criteria.

The shape of the δm sideband is q^2 dependent. For low q^2 the δm sideband is uniform, but for high q^2 it increases with δm, as seen in Figure 4.13. In the q^2 range from 0 to 7.5 $(\mathrm{GeV/c^2})^2$ the sideband is uniform while for higher q^2 the simple counting of events in the sideband overestimates the background. In the q^2 range from 7.5 $(\mathrm{GeV/c^2})^2$ to $q^2 = q^2_{\mathrm{max}} = 10.7$ $(\mathrm{GeV/c^2})^2$ the sideband will overestimate the combinatorial background under the D^* peak. Therefore, events in the δm sideband region with high q^2 are weighted according the caption of Figure 4.13.

4.3.2 Uncorrelated D^*-lepton background

The uncorrelated D^*-lepton background is the background in which a true D^* and a true lepton from different Bs are selected. Since it is required that the D^* and the lepton have different charge the lepton is either from a secondary decay or from an event in which $B^0\bar{B}^0$ mixing occured. Most of the events in this background category come from secondary decays.

This background can be estimated from data using the so-called flipping technique, in which the lepton's momentum direction is reversed. Events passing the event selection criteria with the lepton momentum flipped will be an almost pure sample of background events since the signal events almost always have the lepton and the D^* in oposite hemispheres. Since, by definition, the direction of the D^* and the lepton are uncorrelated this gives a measurement of the amount of this background. There is a small correction to be made for signal events that pass the event selection criteria after the momentum direction of the lepton is reversed. Table 4.2 summarizes the estimated contamination from the uncorrelated background. The background level is higher in the electron channel since there is a lower momentum cut on the electrons than on the muons. Combined in all modes this background is estimated to be 16 ± 6 events.

4.3.3 Continuum

This background is estimated using data collected below the $\Upsilon(4S)$ resonance. In all of the data collected below the $\Upsilon(4S)$ resonance only 2 events passed the selection criteria. Scaling the off-resonance luminosity to the on-resonance luminosity, as given in Table 3.1, this background is estimated to be 5 ± 5 events.

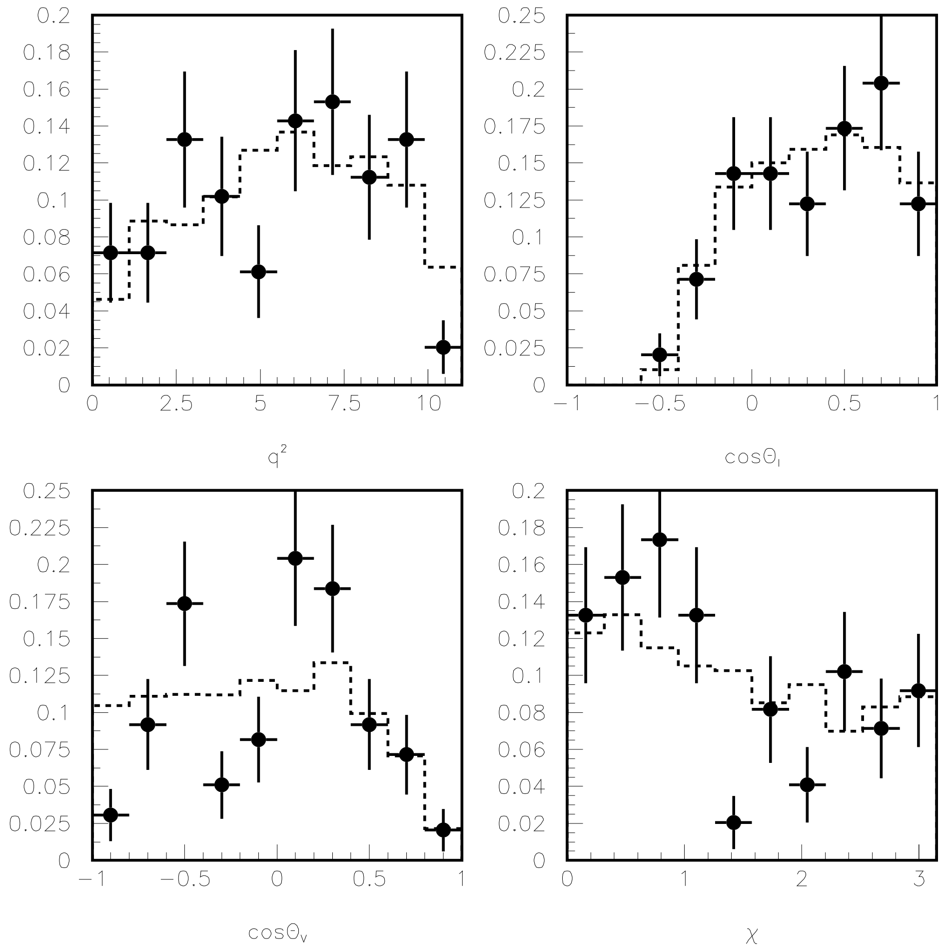

Figure 4.9: These plots shows the δm sideband for $D^0 \to K^-\pi^+$ in the electron channel. The points with error bars are the sideband from the data. The dashed line is a higher statistics sample from TRKSIM.

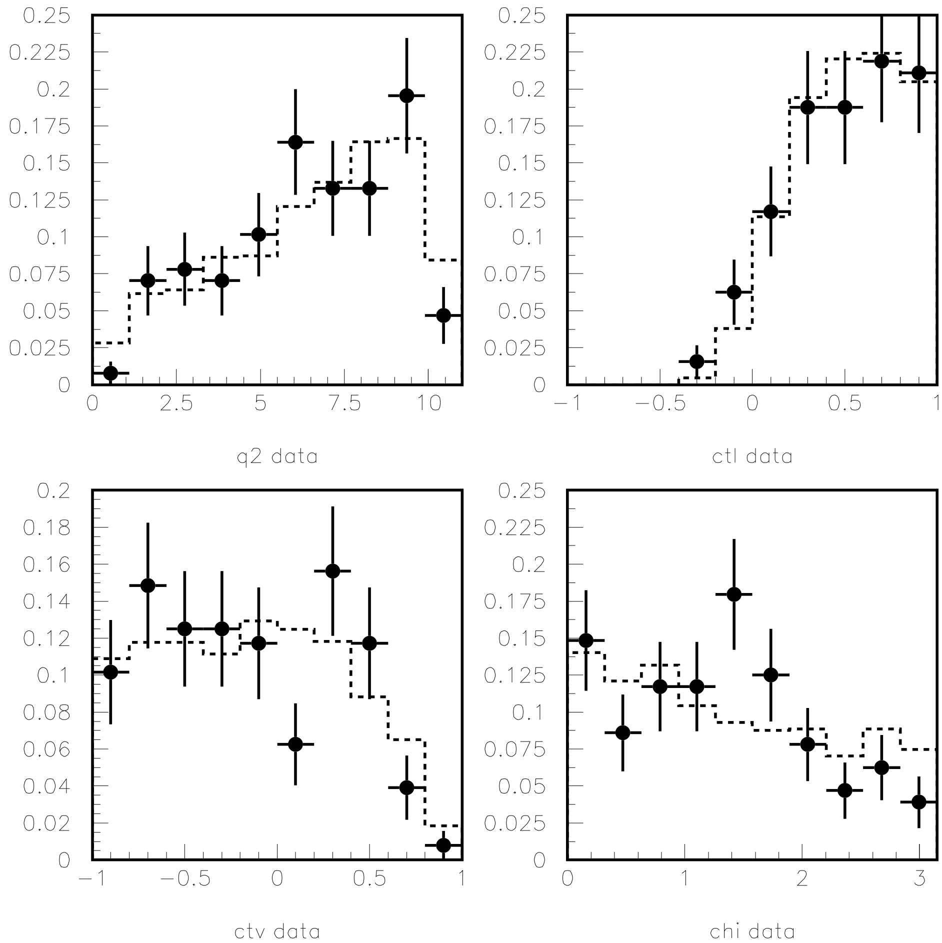

Figure 4.10: These plots shows the δm sideband for $D^0 \to K^-\pi^+$ in the muon channel. The points with error bars are the sideband from the data. The dashed line is a higher statistics sample from TRKSIM.

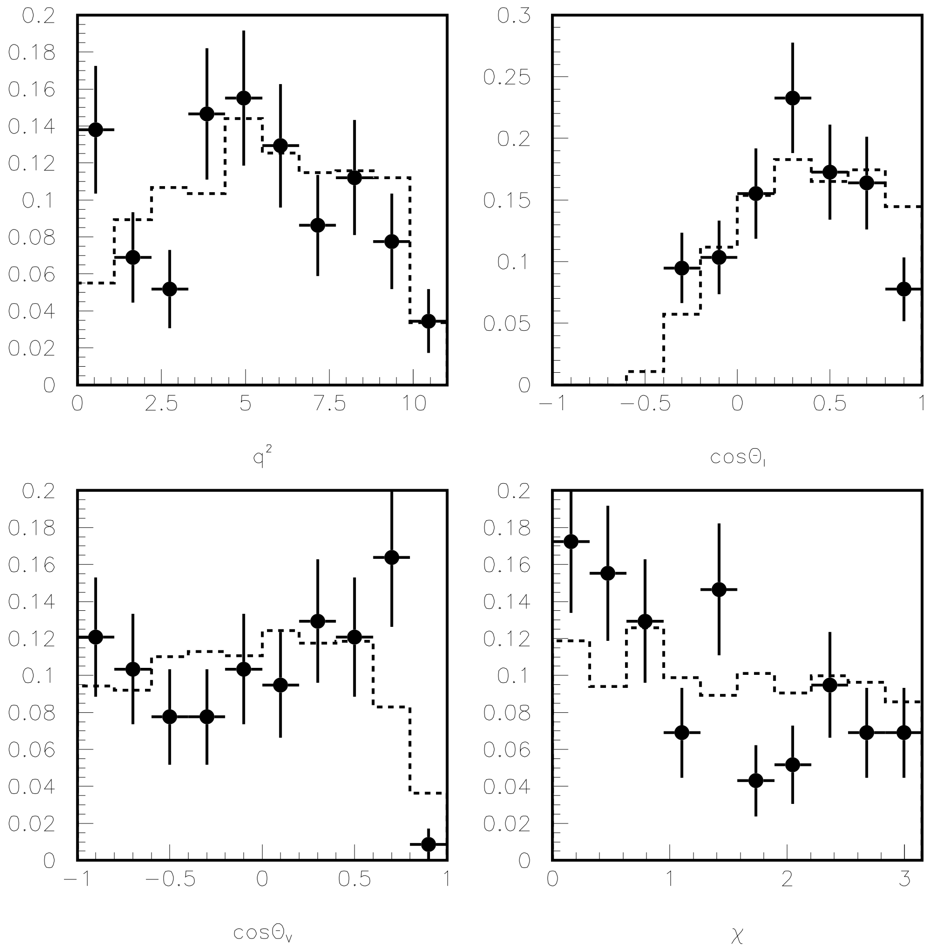

Figure 4.11: The one dimensional projections for the δm sideband in data and TRKSIM in the electron channel with $D^0 \to K^-\pi^+\pi^0$. The agreement between the data sideband events and the TRKSIM events is good.

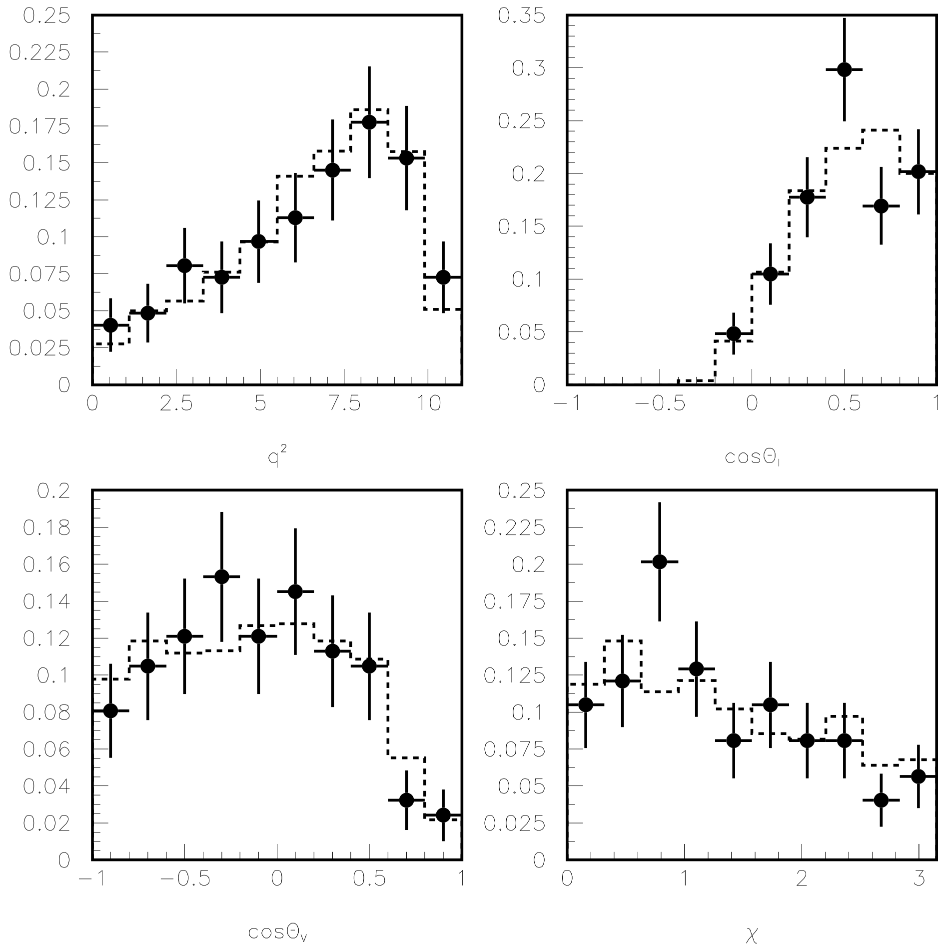

Figure 4.12: The one dimensional projections for the δm sideband in data and TRKSIM in the muon channel with $D^0 \to K^-\pi^+\pi^0$. The agreement between the data sideband events and the TRKSIM events is good.

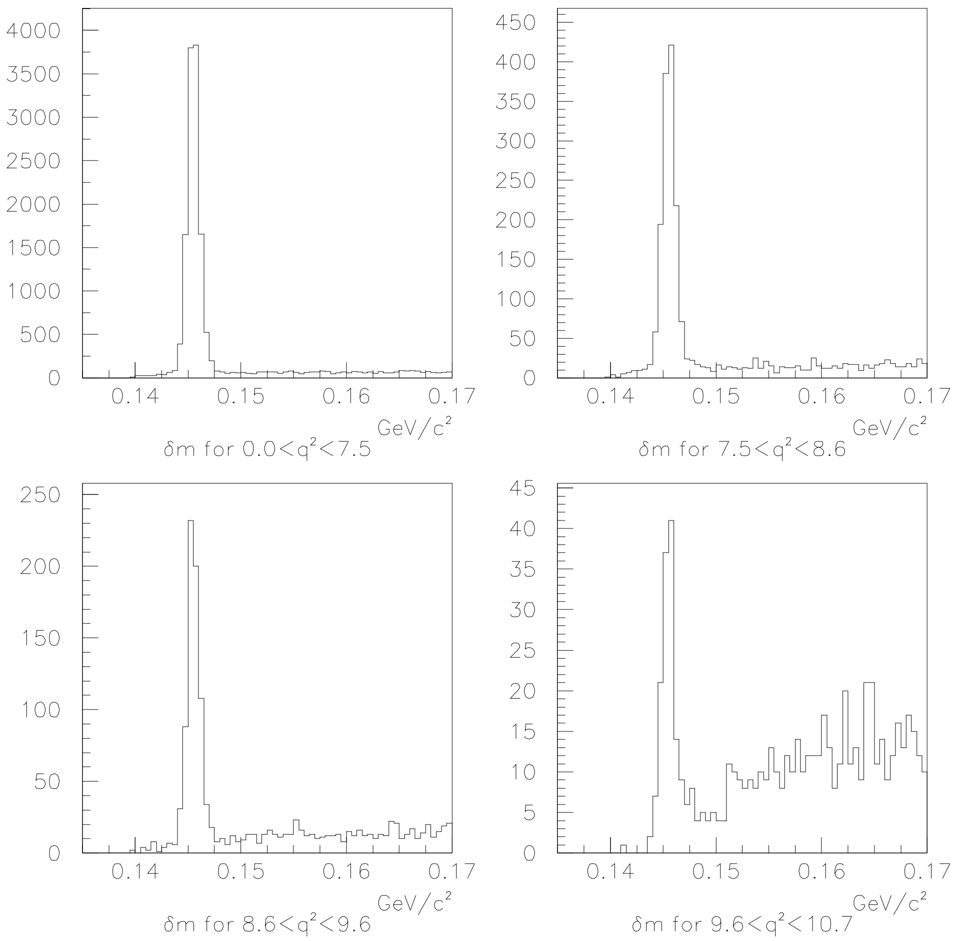

Figure 4.13: These plots show the δm distributions for 4 different ranges of q^2 from TRKSIM Monte Carlo. The first plot shows the δm distribution for q^2 lower than 7.5 (GeV/c^2)^2. Here the sideband region (0.15-017 (GeV/c^2)^2) is flat. For the three other regions the sideband is increasing. In the low q^2 region, the sideband is not de-weighted, i.e., the sideband is taken to be flat over the signal and sideband region. For the second q^2 region the sideband is weighted by 0.7. In the third and the fourth, the highest q^2, region the sideband is weighted by 0.4 and 0.2 respectively. These weights are obtained from fitting the δm distributions.

4.3.4 $B \to D^{**}\ell\nu$ background

This section deals with the contamination from $B \to D^{**}\ell\nu$. D^{**} signifies any excited D or nonresonant $D^{(*)}\pi$ state except the D or the D^*. This gives rise to a contamination of the signal when $D^{**} \to D^{*+}\pi$ and the pion is missed.

As discussed in Section 1.1 the semileptonic width is not saturated by $B \to D\ell\nu$ and $B \to D^*\ell\nu$; these modes only account for about $60 - 70\%$ of the semileptonic width. This indicates that there is a significant semileptonic rate to excited Ds or D^*s. Recent measurements by ALEPH [10], OPAL [11], and CLEO [31] of exclusive B semileptonic decays to D^{**}s provide the first direct measurements of these decays.

These measurements form the basis of the estimate of this background. However, there are many sources of D^{**}s and not all of them have been measured. Therefore the ISGW2 model is used to estimate the amount of some of these D^{**} channels.

The resonant D^{**} states that have been observed are the D_1 and the D_2^*, these states are $L = 1$ (P-wave) mesons. There are four $L = 1$ mesons, whose properties are summarized in Table 4.3. The D_1 and the D_2^* , are explained to be narrow since they decay through a D-wave. The D_1 has $j_l = 3/2$ and is forbidden by HQS to decay in an S-wave, and a P-wave is forbidden by parity. The D_2^* has to decay through a D-wave since S and P-wave are forbidden by spin and parity conservation.

Name	Mass MeV/c^2	Width MeV/c^2	Decays
D_1	2423	18	$D^*\pi$
D_2^*	2458	21	$D^*\pi$, $D\pi$
${}^{1/2}P_0$	2400	150	$D\pi$
${}^{1/2}P_1$	2490	150	$D^*\pi$

Table 4.3: The properties of the $L = 1$, D^{**}, resonances. The two wide states, the ${}^{1/2}P_0$ and the ${}^{1/2}P_1$, have not yet been observed so the mass and widths assigned are estimates.

The other two $L = 1$ resonances, the ${}^{1/2}P_1$ and the ${}^{3/2}P_0$ are allowed to decay in S-wave and are therefore expected to be broad. These states have not yet been observed.

The D_1 decays only to $D^*\pi$ while the D_2^* decays to both $D\pi$ and $D^*\pi$. The rate to $D^*\pi$ is expected to be $\approx 1/2$ of the rate to $D\pi$ for the D_2^*. This is experimentally verified [32] and the ratio is found to be

$$\frac{B(D_2^* \to D\pi)}{B(D_2^* \to D^*\pi)} = 2.3 \pm 0.7. \tag{4.3}$$

The ${}^{1/2}P_1$ decays only to $D^*\pi$ while the ${}^{1/2}P_0$ decays only to $D\pi$, and therefore does not contribute to the background in this analysis.

To estimate the contribution of the various D^{**} states to the background in this analysis, the measurements

$$B(B \to D^{*+}\pi^-\ell\bar{\nu}X) = 0.93 \pm 0.25 \pm 0.18\% \tag{4.4}$$

and

$$B(B \to D_1^0\ell\bar{\nu}X) \times B(D_1^0 \to D^{*+}\pi^-) = 0.51 \pm 0.15 \pm 0.09\% \tag{4.5}$$

from ALEPH [10] are taken as the starting point. The first measurement is inclusive in the sense that it measures all contributions to $B \to D^{*+}\pi^-\ell\bar{\nu}X$ independent of the $D^{*+}\pi^-$ mass, i.e., both resonant and nonresonant. The second measurement measures the fraction that goes through the D_1^0 resonance.

To estimate the contributions from the other two resonant $L = 1$ states, which are sources of background in this analysis, the ISGW2 model is used. The predictions of the ISGW2 model are summarized in Table 4.4. The absolute rate predicted by the model is not used; instead the rate relative to the D_1 is used to predict $B(B \to D_2^*\ell\nu) \times B(D_2^* \to D^*\pi)$ and $B(B \to {}^{1/2}P_1\ell\nu) \times B({}^{1/2}P_1 \to D^*\pi)$. These predictions are summarized in Table 4.5 where the number of events is calculated by

$$\begin{aligned} \#\text{ of events} \;&=\; N_{B\bar{B}} \times Br(B \to D^{**}\ell\nu) \times Br(D^{**} \to D^*\pi) \\ &\quad\times Br(D^* \to D\pi) \times Br(D \to K\pi) \times \epsilon \times 1.5 \end{aligned} \tag{4.6}$$

where $N_{B\bar{B}} = 2.6 \times 10^6$, $Br(D^* \to D\pi) = 0.68$, and $Br(D \to K\pi) = 0.039$. The factor of 1.5 is from isospin to account for the D^{**} production from neutral B-decays.

The last two rows in the Table show the estimated contribution from nonresonant $D^*\pi$. This estimate is obtained by summing the resonant contributions from the exclusive channels and subtracting this from the measured inclusive branching fraction, Eq 4.4. The error on the D_2^* contribution

State	$\frac{\Gamma(B^0 \to D^{**}\ell\bar{\nu})}{\lvert V_{cb}\rvert^2 10^{13}\mathrm{s}^{-1}}$
D_1	0.20
D_2^*	0.10
$^{1/2}P_0$	0.03
$^{1/2}P_1$	0.04

Table 4.4: The prediction of the ISGW2 model for the rates to $L = 1$ D^{**} states.

is 150% of the estimated rate so that it covers the 95% C.L. upper limit from ALEPH. The error on the $^{1/2}P_1$ is taken to be 100%.

The efficiencies for these decays to pass all event selection criteria are obtained from Monte Carlo. Using the form factor predictions from the ISGW2 model the decay chain

$$B^- \to D^{**0}\ell^-\bar{\nu},\ D^{**0} \to D^{*+}\pi^-,\ D^{*+} \to D^0\pi^+ \tag{4.7}$$

is implemented for $D^{**0} = (D_1^0, D_2^{*0}, {}^{1/2}P_1)$. The Monte Carlo correctly treats the correlations among the variables that describes these decay. The decay of the D_1 and the D_2^* to $D^*\pi$ are assumed to be in a D-wave, while the $^{1/2}P_1$ is assumed to go in S-wave. The nonresonant D^{**}s are treated as a wide state with a mass of 3 GeV/c^2.

4.4 Efficiencies and branching fractions

The overall efficiency, from Monte Carlo, is 6.6% for electrons and 7.1% for muons in the $D^0 \to K^-\pi^+$ channel. This gives branching fractions of (5.0 ± 0.4)% for electrons and (4.9 ± 0.4)% for muons. In the $D^0 \to K^-\pi^+\pi^0$ channel the efficiencies are 1.39% and 1.46% for electrons and muons respectively. This gives branching fractions of 4.6 ± 0.7% and 4.3 ± 0.6%. The efficiencies and branching fractions are summarized in Table 4.4. The errors quoted are only statistical; but the efficiencies and branching ratios indicate an agreement with the measured branching fraction [9] for $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$. The background estimates are taken from Table 4.2. Note that the cuts used in this analysis are fairly tight and may not be well modeled by the Monte Carlo. This is however not a problem in this analysis since it is only sensitive to the relative efficiency over the Dalitz plot. The efficiency across

Background mode	$B(B \to D^{**}\ell\nu)\times$ $B(D^{**} \to D^{*+}\pi^-)$ (%)	ϵ (%)	# of events ($K\pi$ channel)
$D_1\ e$	0.51 ± 0.17	1.20 ± 0.11	6.3 ± 2.2
$D_1\ \mu$	0.51 ± 0.17	1.66 ± 0.13	8.8 ± 3.0
$D_2\ e$	0.09 ± 0.13	1.19 ± 0.11	1.1 ± 1.7
$D_2\ \mu$	0.09 ± 0.13	1.94 ± 0.14	1.8 ± 2.7
$1^{1/2}P_1\ e$	0.11 ± 0.11	1.17 ± 0.11	1.3 ± 1.3
$1^{1/2}P_1\ \mu$	0.11 ± 0.11	2.59 ± 0.15	2.9 ± 2.9
Non res. e	0.21 ± 0.42	1.44 ± 0.16	3.1 ± 6.2
Non res. μ	0.21 ± 0.42	1.46 ± 0.16	3.2 ± 6.4

Table 4.5: A model for the contribution of the D^{**} background to the analysis. This is the amount of the background from D^{**}'s that is used to estimate the effect of this background on the fit for the form factor ratios. (For the $K\pi\pi^0$ channel the number of events is just scaled by the yield in the signal, 0.66).

the Dalitz plot is not uniform as seen in Figure 4.14. There are two major effects that causes the efficiency not to be uniform. The minimum lepton energy cut causes the severe loss of efficiency for low $\cos\theta_\ell$. The other major effect comes from the detection efficiency of the slow pion. It causes the loss of efficiency at high q^2 and large $\cos\theta_V$. One of the systematic errors considered is from the slow pion efficiency.

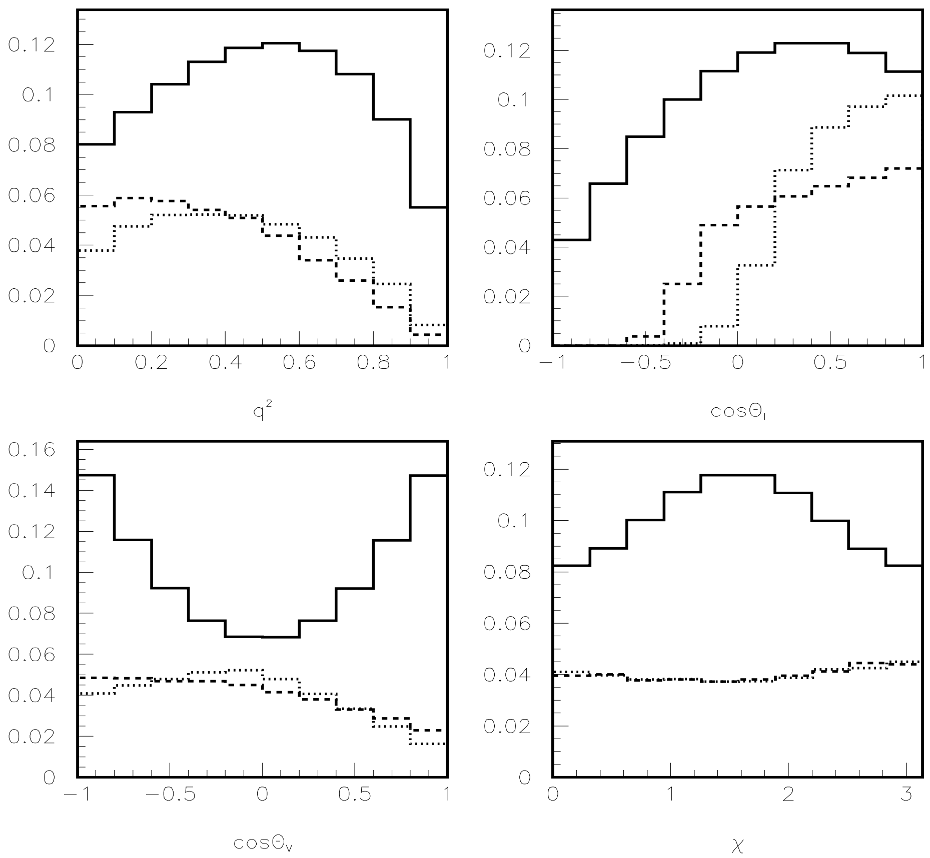

Figure 4.14: Theoretical distributions (solid line) and and detection efficiencies (dashed line for electrons and dotted line for muons). The efficiency is shown on an arbitrary scale. The loss of efficiency at high q^2 and $\cos\theta_V$ is due to the lack of ability to reconstruct the slow pion. The lower lepton momentum cut causes the severe loss of efficiency in the $\cos\theta_\ell$ distribution.

	$K\pi\ e$	$K\pi\ \mu$	$K\pi\pi^0\ e$	$K\pi\pi^0\ \mu$
Yield	228 ± 18	240 ± 20	157 ± 15	152 ± 16
ϵ	6.6	7.1	1.39	1.46
Produced	3454 ± 273	3408 ± 282	11300 ± 1100	10400 ± 1100
Br	5.0 ± 0.4	4.9 ± 0.4	4.6 ± 0.7	4.3 ± 0.6

Table 4.6: The estimated branching fractions in the individual channels. The row with "Produced" is the efficiency corrected yield. Errors quoted are only statistical but there is a good agreement between the individual channels and the previous measurement of $B(\bar{B}^0 \to D^{*+}\ell^-\bar{\nu})$.

Chapter 5

Measurement of the form factors

5.1 Calculation of kinematic variables

For each combination that passes the event selection criteria the kinematic variables q^2, $\cos\theta_\ell$, $\cos\theta_V$, and χ are calculated. It is desired to calculate these variables using the maximal information in order to get the best possible resolution. It is also desirable to get values in the physical range. The fitting method used doesn't require physical values for the kinematic variables, but it is hard to handle the tails of the distributions, i.e., it requires a lot of Monte Carlo events that are expensive to generate. Below, the procedure used to calculate the kinematic variables is described in detail.

This calculation uses two different inertial frames: the laboratory frame, in which all quantities are measured, and the B rest frame, where the different quantities of interest can easily be calculated. Previous work [16] has assumed that the B meson is at rest, i.e., that the laboratory frame and the B meson rest frame are the same. It is found that using the available information on the B momentum helps improve the resolution.

In the laboratory frame the 4-vectors of the D^0, π, and lepton are measured. Using the beam energy constraint the energy and magnitude of the B momentum are also known. This allows calculation of the energy of the missing neutrino and since it is massless it also gives the momentum. As discussed is Section 4.2.4.3 this allows calculation of the cosine of the angle between the B momentum and the momentum of the Y system (the Y system is the D^* and the lepton). This quantity, $\cos\theta_{BY}$, is required to be physical, i.e., $|\cos\theta_{BY}| < 1.0$.

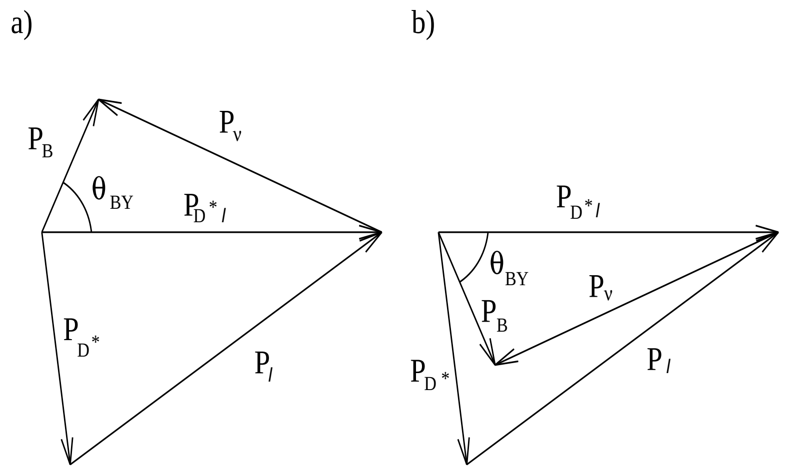

Figure 5.1: The two extreme configurations used to calculate the kinematical variables. The configuration in a) corresponds to minimal q^2 while the configuration in b) is for maximal q^2.

The projection of the B momentum on the Y system is known and there is in fact only one unknown, which is the azimuthal angle of the B momentum around the Y system. The approach taken now is that there are two extreme configurations, one with the B closest to the D^* direction and one with the B direction as far away from the D^* direction as possible, as shown in Figure 5.1. χ is calculated in the first configuration. For both of these configurations the other three variables are calculated and the average is simply taken as the best estimate. In the case of q^2 the result from each of the configurations is restricted to be physical, i.e., from 0 to $q^2_{\max}$.

In the rest frame of the B the following exact relations hold

$$\cos\theta_\ell = \frac{2E_\ell - E_W}{|\vec{P}_{D^*}|}, \tag{5.1}$$

$$\cos\theta_V = \frac{2E_{D^0}M^2_{D^*} - E_{D^*}(M^2_{D^*} + M^2_{D^0} - M^2_\pi)}{|\vec{P}_{D^*}|((M^2_{D^*} - M^2_{D^0} - M^2_\pi)^2 - (2M_{D^0}M_\pi)^2)^{1/2}}, \tag{5.2}$$

$$\cos(\pi - \chi) = \frac{\vec{P}^\perp_\pi \cdot \vec{P}^\perp_\ell}{|\vec{P}^\perp_\pi||\vec{P}^\perp_\ell|}, \tag{5.3}$$

$$\vec{P}^\perp_\pi = \vec{P}_\pi - \frac{\vec{P}_\pi \cdot \vec{P}_{D^*}}{|\vec{P}_{D^*}|^2}\vec{P}_{D^*}, \tag{5.4}$$

$$\vec{P}_\ell^{\perp} \;=\; \vec{P}_\ell - \frac{\vec{P}_\ell \cdot \vec{P}_{D^*}}{|\vec{P}_{D^*}|^2}\vec{P}_{D^*}. \tag{5.5}$$

The symbol $^{\perp}$ means perpendicular to the D^* direction, and the masses M_{D^*} and M_{D^0} are the invariant masses of the $K\pi\pi$ and $K\pi$ or $K\pi\pi\pi$ and $K\pi\pi$ combinations. Using Lorentz invariance, all quantities in the formulas above can be expressed in laboratory frame quantities (which are different from the B rest frame). For example, the energy of the lepton in the B rest frame can be written as

$$E_\ell^{(B)} = p_B^\mu p_{\ell\mu}/M_B \tag{5.6}$$

and the W energy can be expressed as

$$E_W^{(B)} = p_B^\mu q_\mu/M_B. \tag{5.7}$$

In the same way all other energies needed are calculated. The expression for $\cos(\pi - \chi)$ is expanded in terms of scalar products of 3-vectors. These products are evaluated in the B frame, for example

$$\vec{P}_\ell \cdot \vec{P}_{D^*} = E_\ell E_{D^*} - p_\ell^\mu p_{\mu D^*}. \tag{5.8}$$

Using similar relations, all of the needed scalar products can be evaluated in the B rest frame. The reason for calculating $\pi - \chi$ is that it is numerically easier to use the slow pion since it has a relatively larger momentum component perpendicular to the D^* momentum than the D^0.

Figure 5.2 shows the resolutions obtained for $\cos\theta_\ell$ in four different q^2 ranges. The plots show the generated value for $\cos\theta_\ell$ versus the reconstructed value. The resolution is best for low q^2 which corresponds to maximum D^* momentum in the B frame. At high q^2, i.e., at low D^* momentum the resolution is worse since the impact of the unknown B direction is bigger. The resolution in $\cos\theta_\ell$ was found to be between 0.044 and 0.067 for increasing q^2. In Figure 5.3 the same plots are shown for $\cos\theta_V$. In the same way it is found here that the resolution gets worse with increasing q^2. For low q^2 the resolution is found to be 0.043 while for high q^2 the resolution is 0.092. In Figure 5.4 the resolution of the q^2 and χ variables are shown. The fairly long tails on the χ distribution comes largely from combinations with $\cos\theta_\ell$ or $\cos\theta_V$ close to ± 1, which corresponds to configurations where the χ angle becomes undefined. Note that all terms in the decay rate that contain χ are multiplied by at least one factor each of $\sin\theta_\ell$ and $\sin\theta_V$.

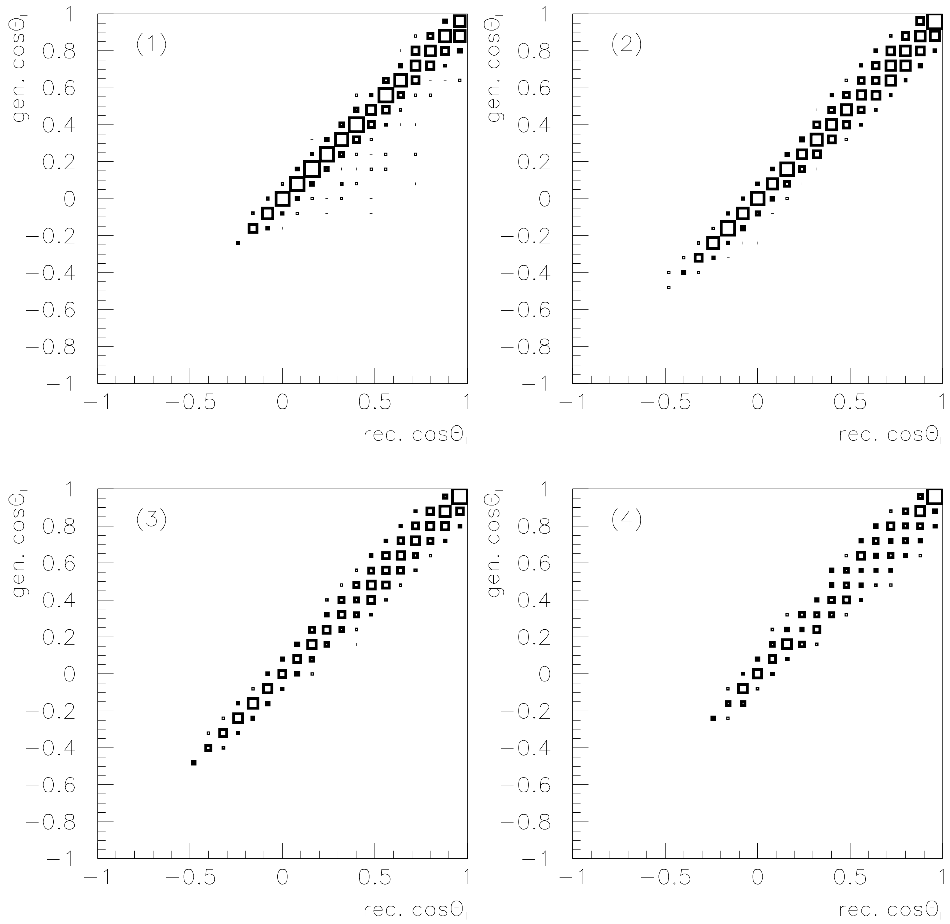

Figure 5.2: The reconstructed v.s. generated values of $\cos\theta_\ell$ in the four q^2 ranges: (1) $q^2 < 2.5$ GeV^2/c^4 (2) 2.5 GeV^2/c^4 $< q^2 <$ 5.0 GeV^2/c^4 (3) 5.0 GeV^2/c^4 $< q^2 <$ 8.0 GeV^2/c^4 (4) 8.0 GeV^2/c^4 $< q^2$

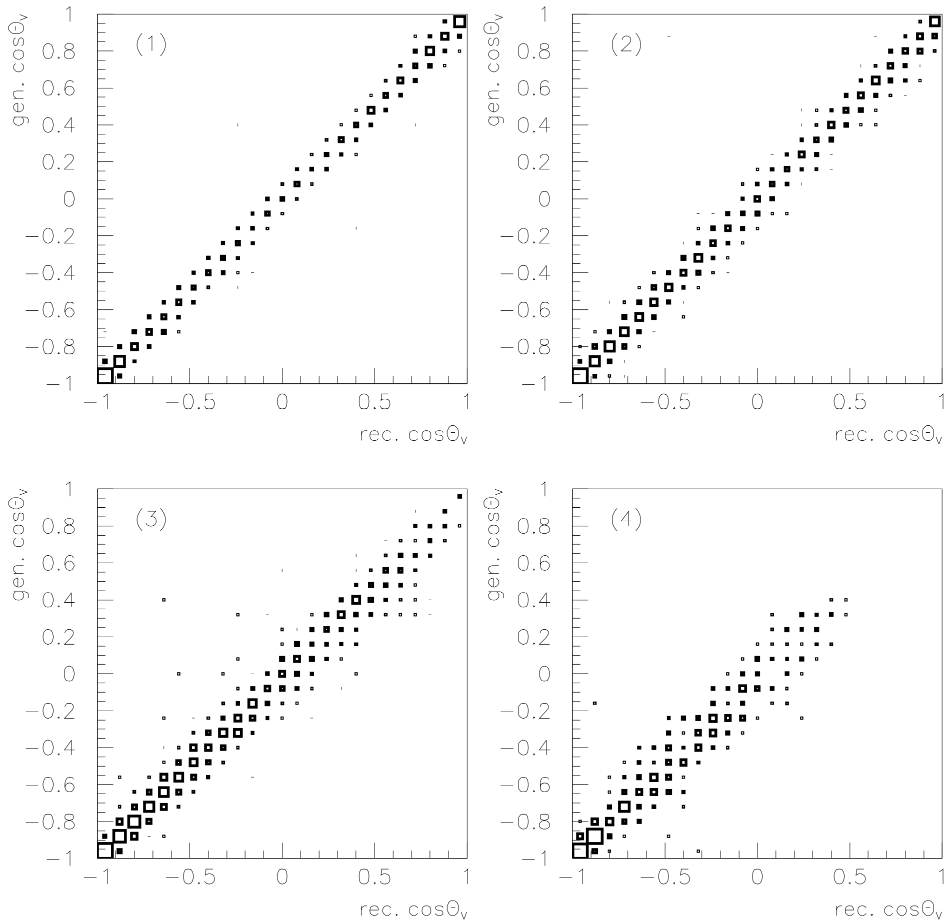

Figure 5.3: The reconstructed v.s. generated values of $\cos\theta_V$ in the four q^2 ranges: (1) $q^2 < 2.5$ GeV^2/c^4 (2) 2.5 GeV^2/c^4 $< q^2 <$ 5.0 GeV^2/c^4 (3) 5.0 GeV^2/c^4 $< q^2 <$ 8.0 GeV^2/c^4 (4) 8.0 GeV^2/c^4 $< q^2$

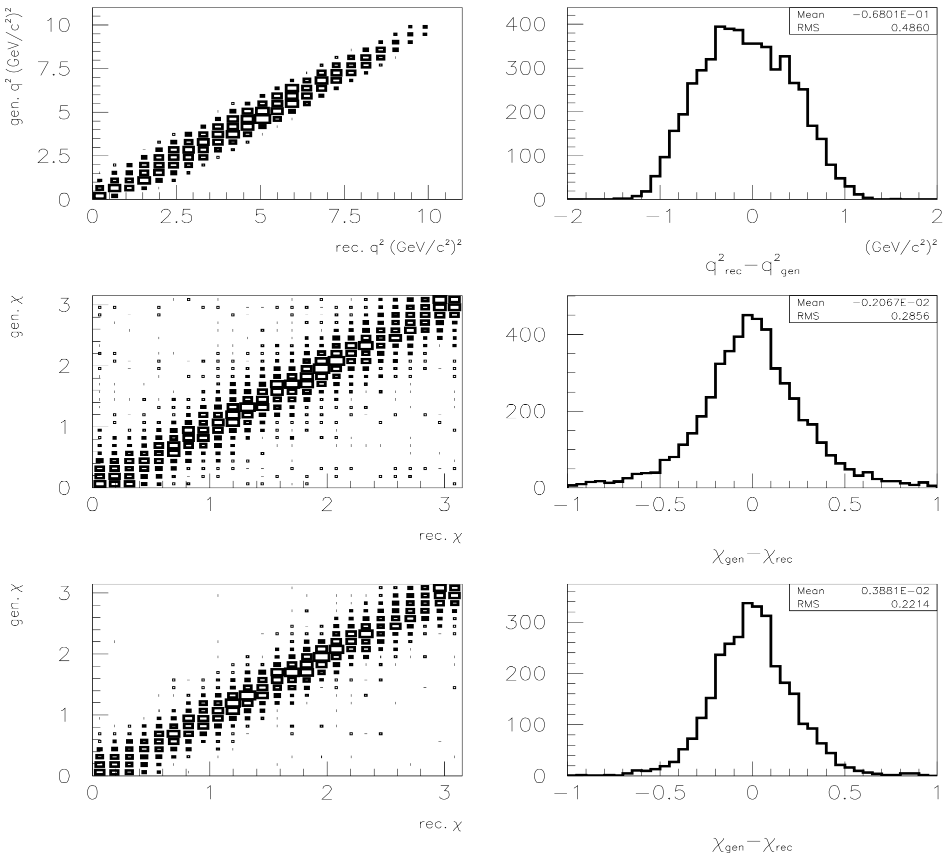

Figure 5.4: The resolution on q^2 and χ. The resolutions on q^2 is about 0.5 $(\mathrm{GeV/c^2})^2$. The resolution on χ is about 0.29 for all events. The second plot of χ is for events with $\sin\theta_V$ and $\sin\theta_\ell$ greater than 0.2. This gives a better resolution (0.22) since χ becomes undefined when either $\sin\theta_V$ or $\sin\theta_\ell$ goes to zero. Note that all terms in the decay rate that contain χ vanish as $\sin\theta_V$ or $\sin\theta_\ell$ goes to zero.

5.2 Fitting method

In this section the fitting method used in this analysis is discussed. The first part of this section will describe how Monte Carlo events and re-weighting is used to calculate the likelihood. The second part will describe how backgrounds are treated in the fit.

5.2.1 Maximum likelihood method with Monte Carlo weighting

The data need to be fitted to a theoretical decay distribution $\Gamma(x, \mu)$ given by Eq. 2.73. Here $x = (q^2, \cos\theta_\ell, \cos\theta_V, \chi)$ represents the observed quantities, and $\mu = (R_1, R_2, \rho^2_{A_1})$ represents the parameters that specify the theory. These are the quantities that the fit will determine. Unfortunately, the observed distribution is not $\Gamma(x, \mu)$. What is observed is a smeared and acceptance affected distribution, $F(\tilde{x}, \mu)$

$$F(\tilde{x}, \mu) = \int \mathcal{D}(\tilde{x}, x)\Gamma(x, \mu)dx. \tag{5.9}$$

The function $\mathcal{D}(\tilde{x}, x)$ represents the effects of smearing and acceptance. There are several different sources of smearing in this experiment. There is the detector resolution which gives a negligible smearing in this analysis. The more severe source of smearing comes from the partially unknown B motion. The effects of acceptance are also very large in this measurement since, for example, the slow π in the decay $D^{*+} \to D^0\pi^+$ cannot be observed when the D^{*+} is at rest. This gives an efficiency near zero at $q^2_{\max}$.

A method for doing these fits that incorporates the effects of acceptance and smearing is needed. One way of doing this was pioneered by the E691 experiment in their measurement of the $D \to K^*\ell\nu$ semileptonic form factors [33, 34]. It is an unbinned likelihood fit that uses Monte Carlo events to numerically evaluate the likelihood. It uses a re-weighting technique to evaluate the likelihood for different set of parameters μ. The idea behind this method is described below.

Monte Carlo events are generated according to some theory, specified by the parameters $\hat{\mu}$. The events are passed through a full Monte Carlo and reconstructed applying the same selection criteria as for the data. The reconstructed events are distributed according to

$$F(\tilde{y}, \hat{\mu}) = \int \mathcal{D}(\tilde{y}, y)\Gamma(y, \hat{\mu})dy. \tag{5.10}$$

The observables associated with Monte Carlo events are denoted by y and those for the data is denoted by x. The symbols $\tilde{x}$ and $\tilde{y}$ refer to the reconstructed (observed) value of the variables. For the Monte Carlo events both y and $\tilde{y}$ are known while for the data only $\tilde{x}$ is known. A maximum likelihood fit is performed so

$$ln(\mathcal{L}(\mu)) = \sum_{i=1}^{n}(ln(F(\tilde{x}_i, \mu)) - ln(N(\mu))) \tag{5.11}$$

needs to be evaluated. Here, $N(\mu) = \int F(\tilde{x}, \mu))d\tilde{x}$ is the normalization, and the index i runs over all events in the data sample. First, note that if the Monte Carlo is generated according to $\hat{\mu}$ it is easy to evaluate the likelihood for that theory. Since the $\tilde{y}$ are distributed according to $F(\tilde{y}, \hat{\mu})$ by simply selecting a volume V_j centered at each observed data point, $\tilde{x}_j$, in the space of $\tilde{x}$ and count how many Monte Carlo events fall, within V_j. Thus,

$$\begin{gathered}\frac{1}{N}(\frac{dN}{d\tilde{x}}(\tilde{x} = \tilde{x}_i))|_{\mu=\hat{\mu}} = \frac{F(\tilde{x}_i, \hat{\mu})}{N(\hat{\mu})} = \\ \frac{Number\ of\ \{\tilde{y}_i\}\ in\ V_i}{V_i \times (Number\ of\ reconstructed\ MC\ events)}.\end{gathered} \tag{5.12}$$

The trick is to figure out how to evaluate the likelihood for another theory, i.e., for a general μ not equal to $\hat{\mu}$. This can be done by assigning the weight $W(y_i, \mu) = \Gamma(y_i, \mu)/\Gamma(y_i, \hat{\mu})$ to each Monte Carlo event $(y_i, \tilde{y}_i)$. The Monte Carlo events with the weight $W(y_i, \mu)$ are distributed according to $F(\tilde{x}_j, \mu)$. The main point is that the effects of the acceptance and the smearing for any parameter set μ can be evaluated simply by re-weighting Monte Carlo events distributed according to a parameter set $\hat{\mu}$, as long as the set $\hat{\mu}$ populates all of phase space.

For a set of n data events and m Monte Carlo events the log-likelihood can be written as

$$ln(\mathcal{L}(\mu)) = \sum_{i=1}^{n} ln\left[\frac{\sum_{\tilde{y}_j\ in\ V_i} W(y_j, \mu)}{C(\mu)V_i}\right] \tag{5.13}$$

where V_i is the volume centered around $\tilde{x}_i$ and $C(\mu) = \sum_{j=1}^{m} W(y_j, \mu)$ is the normalization of the weighted MC sample. When $\mu = \hat{\mu}$ the expression simplifies to the case above.

5.2.2 Implementation of the fitting method

The volume around each data point in which the density of Monte Carlo events is computed will be called the search volume. This volume is a 4-

dimensional box with half sides $\Delta y = \Delta q^2/q^2_{\text{max}}$, $\Delta\cos\theta_\ell$, $\Delta\cos\theta_V$, and $\Delta\chi$. When a data point falls so close to the edge of phase space that the distance between the edge and the point is smaller than the half length of the side of the search volume, the half length is taken to be the distance between the point and the edge. It is preferable to have a search volume that is symmetric around the data point. However, the way it was described above will lead to very small boxes when the data point falls very close to the edge. So for points very close to the boundary the requirement of symmetric search volumes are relaxed. For a point that falls closer to the edge than one-half of the specified half side, the half side of the volume that is towards the interior of the space is taken to be one half of the specified half side, see Figure 5.6.

This is done so that the Monte Carlo sample needed to sample a data point close to the boundary doesn't need to be infinitely large. This introduces possible systematic effects that are discussed in Section 6.1.

5.2.3 Tests of fitting method

Tests have been performed to prove that the method works and that the errors that are returned from the fit are correct. The first test is to show that the fit can reconstruct the correct central values. Fits are done to different Monte Carlo samples generated with different values of R_1, R_2, and $\rho^2_{A_1}$. This test is done on Monte Carlo samples from TRKSIM, since high statistics are needed so that problems in the fitting method are not hidden by large statistical errors. To calculate the likelihood function, approximately 100,000 accepted TRKSIM events where used. The samples that the fits where done on had approximately 10000 events each. The results of these fits shows that the fitting method is able to extract the parameters. There is, however, a small bias, which is discussed in Section 6.1.1.

The purpose of the next test is to verify that the errors from the fit are correct. Because a likelihood method is used, the 1σ error corresponds to a change of 1/2 in the $ln(\mathcal{L})$. To test the gaussian nature of the errors 20 Monte Carlo samples were fit and the widths of the distributions of R_1 and R_2 from these fits were in agreement with the errors from the fits. In Figure 5.5 the result of the fits for R_1 and R_2 are shown together with a fit to a gaussian. This shows that the errors obtained from the fit agrees with the spread on the measured quantities.

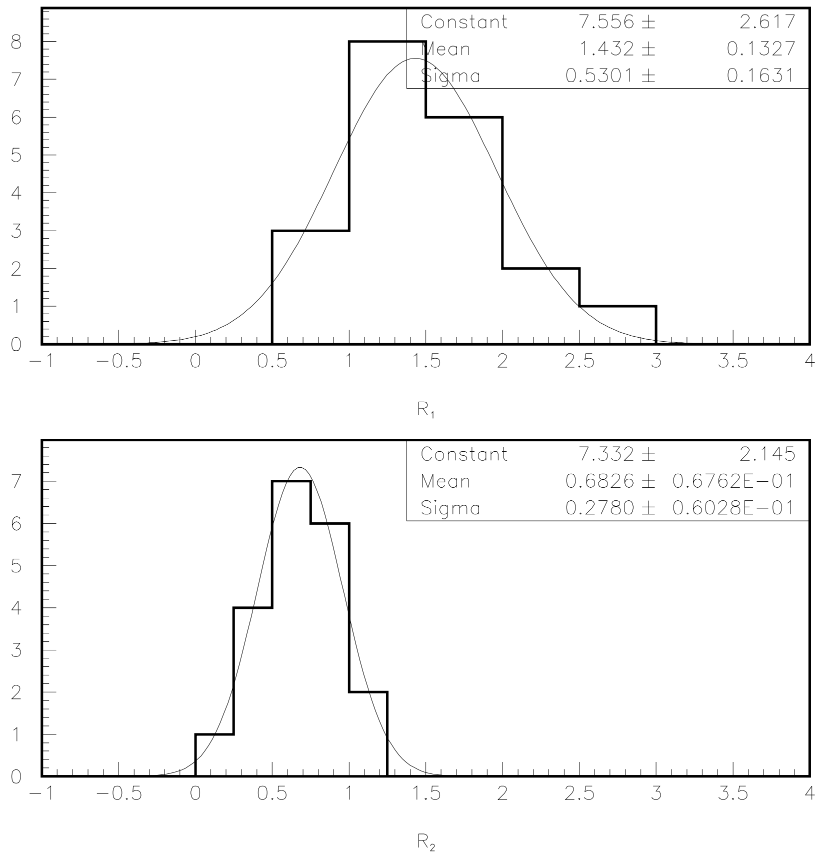

Figure 5.5: The results of each fit for R_1 and R_2 are put into these histogram and a fit is performed to determine the spread of the distributions. For the fit for R_1, a σ of 0.53 was found. This agrees well with the average error of 0.50 for the fits. For R_2 a sigma of 0.28 was found, compared to the average error of 0.23 from the fits.

5.2.4 Backgrounds in the fit

The method used to deal with the backgrounds is a standard likelihood fit to a signal and a background shape. In the fits that are done in this analysis the amount of background is fixed. The fake D^* and the uncorrelated D^*-lepton backgrounds are incorporated in the fit.

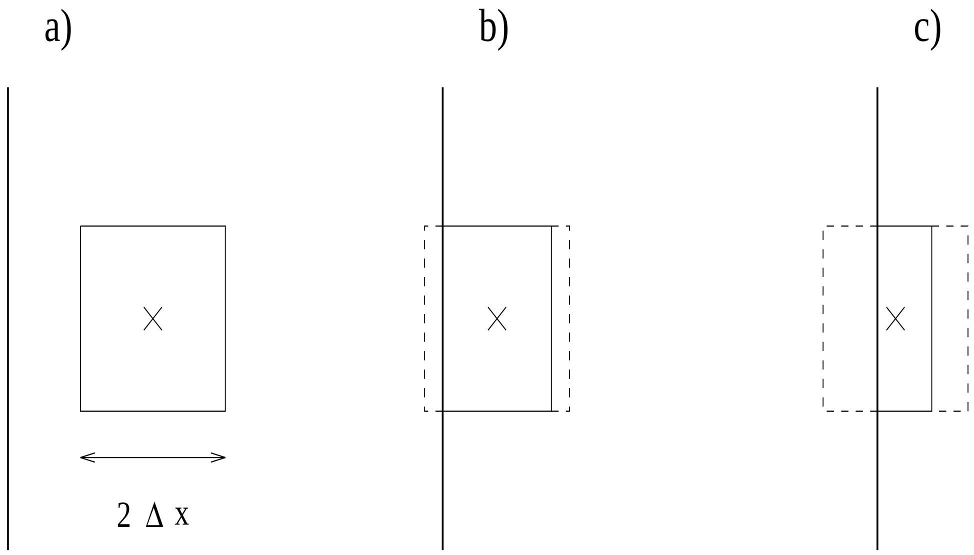

Figure 5.6: The three figures shows the three different cases for how the search volumes are assigned. The first case, a), is when the data-point is so far from the edge of phase space that the full volume can be taken. As the data point gets close to the edge of phase space, b), the half sides are made shorter so that the volume remains symmetric around the data point. The requirement that the search volume is symmetric around the data point is relaxed when the point is very close to the edge, c).

5.3 The Cramér-Rao theorem

5.3.1 Definition of Information and estimates of the variance matrix

It is important to understand the errors obtained from the form factor fits. The motivations are several. First, the size of the errors obtained from the fit needs to be understood, there are many ways of verifying this (see, for example, Section 5.2.3), of which the Cramér-Rao theorem provides one test. Another point that is probably more important is that it helps understanding the measurement better. Below, a set of plots showing the information density as a function of the variables used in the analysis are shown. These plots show which parts of phase-space contributes information to the measurement and provides valuable information for deciding what the optimal event selection criteria are.

The tool for doing this comes from the Cramér-Rao theorem, which allows one to calculate a lower bound on the variance for parameters estimated from N events distributed according to a probability distribution $P(x, \mu)$.

The definition of information according to R.A Fischer [35], for N events is given by

$$I(\mu) = N \int \left[\frac{\partial P(x, \mu)}{\partial \mu} \right]^2 \frac{1}{P(x, \mu)} dx. \tag{5.14}$$

This gives the information on one parameter μ and $P(x, \mu)$ is a normalized probability distribution

$$\int P(x, \mu) dx = 1. \tag{5.15}$$

When there are several parameters $\mu = (\mu_1, \mu_2 ... \mu_n)$ the definition generalizes in the following way and the information becomes an $n \times n$ matrix

$$I_{ij}(\mu) = N \int \left[\frac{\partial P(x, \mu)}{\partial \mu_i} \frac{\partial P(x, \mu)}{\partial \mu_j} \right] \frac{1}{P(x, \mu)} dx. \tag{5.16}$$

The information matrix is simply related to an estimate of the error matrix. The matrix $V = I^{-1}$ gives lower bounds for the errors of an estimate of the parameters μ. The information is an additive quantity and can be thought of as an estimate for the weight matrix.

5.3.2 Applications to $B \to D^* \ell \nu$

Table 5.1 gives the minimal errors for N events according to the Cramér-Rao theorem from a fit to q^2, $\cos \theta_\ell$, $\cos \theta_V$, and χ. The errors are for the case when there is no smearing or acceptance effects, which corresponds to doing a fit to generator level Monte Carlo.

The effects of acceptance on the fit variables are significant and cannot be neglected. Using Monte Carlo, the efficiency $\epsilon(x)$, where $x = (q^2, \cos \theta_\ell, \cos \theta_V, \chi)$, is determined. Using $P(x, \mu) = \epsilon(x) \Gamma(x, \mu)$ the errors on a distribution that

N	$\sigma_{R_1}^{\min}$	$\sigma_{R_2}^{\min}$	$\sigma_{\rho^2}^{\min}$
100	0.35	0.36	0.35
500	0.16	0.16	0.16
1000	0.11	0.12	0.11

Table 5.1: The error bounds obtained from Cramer-Rao's theorem for a fit of generator level distributions. The errors are evaluated at values of R_1, R_2, and $\rho^2_{A_1}$ obtained by the measurement presented in this dissertation.

is affected by acceptance are estimated. This does not include effects of smearing. Note that if the part of phase space that contains little information on a parameter has low efficiency, then the error on that parameter can be smaller than that obtained from the same number of events with no acceptance loss.

Using the acceptance for electrons and muons separately and combining 385 electrons and 394 muons, as was found in the data, the variance matrix for the measurement of the form factors is estimated to be

$$V = \begin{bmatrix} 0.056 & -0.029 & 0.017 \\ -0.029 & 0.025 & -0.017 \\ 0.017 & -0.017 & 0.018 \end{bmatrix}. \tag{5.17}$$

This estimate of the errors neglects two effects; the smearing of the variables and the background in the fit. The bounds on the errors from this estimate is 0.24, 0.16, and 0.13 for R_1, R_2, and $\rho^2_{A_1}$ respectively. The estimated error on $\rho^2_{A_1} = 0.13$ is closest to the actual error from the fit, this is most likely due to the fact that $\rho^2_{A_1}$ is mainly determined from the q^2 distribution, which is not severely affected by smearing.

The variance matrix estimate above gives a quite good estimate of the correlation coefficients: $\rho_{R_1R_2} = -0.78$, $\rho_{R_1\rho^2} = 0.53$, and $\rho_{R_2\rho^2} = -0.80$.

In Figure 5.7 a sample of one dimensional information plots are shown. The plots shows the information density for the theoretical distributions, or what would have been observed in an experiment with a perfect detector. The plots in the first column show the information on R_1 as a function of the three variables q^2, $\cos\theta_\ell$, and $\cos\theta_V$. In particular in the second plot the information as a function of $\cos\theta_\ell$ is interesting. As discussed earlier R_1 controls the forward-backward asymmetry of the lepton in the W rest frame. Therefore, the information on R_1 is largely coming from events that measure the forward-backward asymmetry, i.e., events close to $\cos\theta_\ell = \pm 1$. The plots in the second column show the information density of R_2. R_2 controls the polarization of the D^*, and the concentration of information at low q^2 is understood since here the D^* is purely longitudinal.

In Figure 5.8 the same plots are shown, but the plots now include the effects of acceptance. In particular, the information density as a function of $\cos\theta_\ell$ is affected. The minimum lepton energy cut rejected the region of $\cos\theta_\ell$ that provided the most information on R_1. Even though the acceptance is falling steeply for $\cos\theta_\ell$ below zero, the events there provides a lot of information in determining R_1. This is also seen in the subtraction of the uncorrelated background where just a few events at low $\cos\theta_\ell$ has a large impact on the measurement of R_1.

5.4 Monte Carlo generation

This analysis depends heavily on Monte Carlo simulation to model the signal and background shapes. The standard CLEO II Monte Carlo, CLEOG, is based on the GEANT [36] detector and physics simulation package from CERN. The CLEOG Monte Carlo performs a detailed simulation of all detector elements and physics processes. The main problem with this Monte Carlo generation is that it is very CPU intensive. This has been a problem in this analysis in particular since the amount of Monte Carlo needed in the signal modes to model the acceptance and smearing in the fit is large. Approximately 6×10^6 events have been generated. This corresponds to approximately 300 CPU days on a DEC Alpha 600. The University of Minnesota and the University of Hawaii group's help in generating the large Monte Carlo samples is greatly appreciated. The Monte Carlo samples for the signal channel used in the fits for the form factors are generated according to the form factor predictions of the ISGW [18] model.

In modeling the fake D^* background, discussed in Section 4.3.1, the complete detector simulation, using CLEOG, is too slow since the statistics needed corresponds to hundreds of millions generated events. Here a fast Monte Carlo, TRKSIM, is used. This fast Monte Carlo does not perform a full simulation of the detector; instead it smears the generated momenta of the particles and convolutes these parameters with an acceptance function.

The fast Monte Carlo is used to model combinatorics, i.e., it is used to model distributions that are not sensitive to the details in the simulation, like effects of acceptance and smearing. Note also that only the shapes are

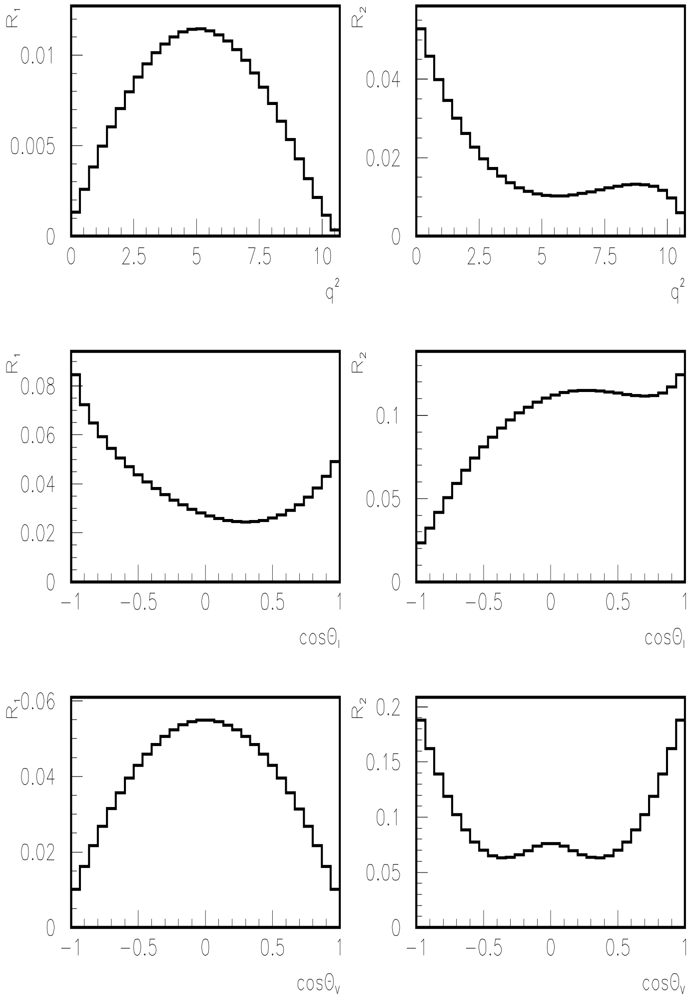

Figure 5.7: The four dimensional information on R_1 (left) and R_2 (right) in the variables q^2, $\cos\theta_\ell$ and $\cos\theta_\ell$ for the theoretical distribution.

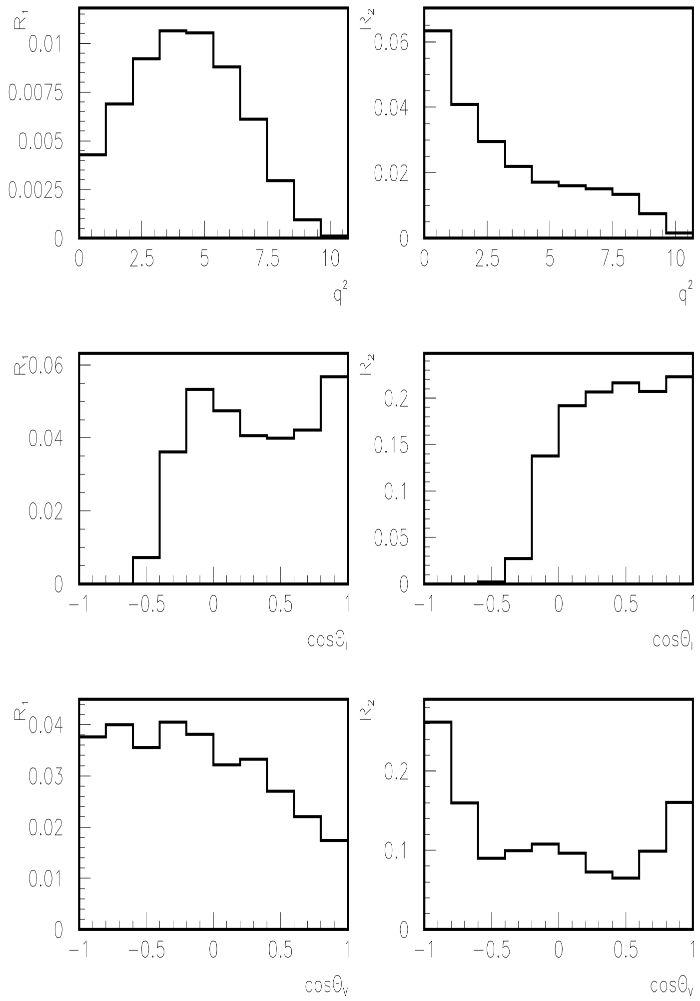

Figure 5.8: The four dimensional information on R_1 (left) and R_2 (right) in the variables q^2, $\cos\theta_V$ and $\cos\theta_\ell$ for the observed distribution.

used from TRKSIM, not the absolute normalization.

In this dissertation the term Monte Carlo refers to the full CLEOG simulation except if otherwise is explicitly stated.

Chapter 6

Systematic studies

The systematic uncertainties come from several different sources. The first systematic discussed is related to the fitting method. Here the effects of the finite volume sizes and the finite Monte Carlo statistics used in the fits are discussed. Then the systematic effects related to the backgrounds, the fake D^*s and the $B \to D^{**}\ell\nu$ are discussed in detail. There is also a discussion of the uncorrelated background. The third type of systematic errors considered are related to the Monte Carlo modeling of efficiencies. Here the primary interest is in the slow pion efficiency, but also the Monte Carlo modeling of the lepton efficiency and the efficiency of the δm cut is considered.

Since the finite volume sizes leads to a bias in the estimated parameters, a correction is determined that is applied to the fit of the data. Similarly, there is a correction from omitting the D^{**} background in the fit.

6.1 Fitting method

The systematic related to the fitting method is described here. The most important effect is the finite volume sizes. The study of this effect leads to a set of corrections that are applied on the data.

6.1.1 Finite volume sizes

The goal of this section is to show that the effects of the finite volume sizes are understood as a function of the result from the form factor fit. This will give a set of corrections that will be applied to the fit to correct for the effects of the finite volume sizes. Again, the reader is reminded that the fit is not a binned likelihood fit. Instead, Monte Carlo events are used to numerically calculate the likelihood. This calculation involves selecting Monte Carlo events around each data point in a finite volume as discussed in Section 5.2.2.

In some cases, the effects of a finite volume size can be qualitatively understood from simple considerations. In this analysis, the effect of a finite volume size in $\cos\theta_V$ is relatively easy to understand. The distribution in $\cos\theta_V$ measures the polarization of the D^*, which in turn is governed by the form factor ratio R_2. A smaller R_2 corresponds to a larger longitudinal polarization and a $\cos\theta_V$ distribution that has a large $\cos^2\theta_V$ component, i.e., one that is peaked at ± 1. Similarly, larger values of R_2 lead to a distribution that is more uniform in $\cos\theta_V$. The fit uses a sampling technique to evaluate the likelihood for each data point. The effect of sampling in a finite volume is that the probability is evaluated as an average around the point. This means that the probability function is evaluated according to some effective distribution that is a smeared–out version of the true distribution. This has the effect that the likelihood, in the case of finite $\cos\theta_V$ volume, is evaluated according to a function with a *larger* R_2. As a result, the best fit will systematically estimate too *small* a value for R_2. This is indeed what is observed, as shown in Figure 6.2.

The discussion will now focus on how these effects are quantitatively studied and how a correction is derived that will be applied on the data. There are two steps in deriving this correction. The first part is a fast Monte Carlo (TRKSIM) study that shows that, for the volume sizes that are used in this analysis, there are indeed systematic effects that are due to the finite volume sizes. The purpose of this first step is to establish that this is a volume size effect and not something else. The next step is to evaluate a correction that will be applied to the data. There are several options for how this might be done. The simplest is just to get a correction from the TRKSIM study and apply this to the data. A little more elaborate method of doing essentially the same thing is to generate a sample of CLEOG events according to the values obtained from the fit of the data and then fit this sample to get the effect of the finite volume size. The corrections that are applied to the data will be taken from a fit of Monte Carlo events that have the complete detector simulation. However, the sample is not generated with the form factors that were obtained in the fit. Instead, a sample is drawn from the large Monte Carlo sample, which is used for the form factor fits, that is distributed according to the measured form factors in the data. This is not exactly an optimal procedure since the sample that is fitted is a subsample of the Monte Carlo sample used in the fit. However, by dividing the fit into two independent subsamples it is shown that if there is a bias, it is very small compared to the size of the effect under study.

TRKSIM samples are generated according to the HQET distribution for a variety of different values of R_1, R_2, and $\rho^2_{A_1}$. These samples are then fit, using a sample generated according to the ISGW model. In Figures 6.1–6.3 the generated versus fitted values are shown for R_1, R_2, and $\rho^2_{A_1}$. It is clear that there is a bias towards underestimating the value of R_2 in the fit. This is a noticeable effect in these high statistics fits. The statistics in these fits are about 20 times higher than in the fit of the data.

The first step is to establish that this indeed is a volume size effect. The fitted parameters, $\eta_i^{\rm fit}$ ($\eta_i^{\rm fit} = R_1,\ R_2$, or $\rho^2_{A_1}$) are functions of the search volumes

$$\eta_i^{\rm fit} = \eta_i^{\rm true} - \delta\eta_i(\Delta q^2, \Delta\cos\theta_\ell, \Delta\cos\theta_V, \Delta\chi). \tag{6.1}$$

$\delta\eta_i$ is the correction that needs to be determined. To proceed, the correction is expanded in the finite volume sizes,

$$\begin{aligned}
&\delta\eta_i(\Delta q^2, \Delta\cos\theta_\ell, \Delta\cos\theta_V, \Delta\chi) = \\
&\qquad \frac{\partial\delta\eta_i}{\partial\Delta q^2}\Delta q^2 + \frac{\partial\delta\eta_i}{\partial\Delta\cos\theta_\ell}\Delta\cos\theta_\ell + \\
&\qquad \frac{\partial\delta\eta_i}{\partial\Delta\cos\theta_V}\Delta\cos\theta_V + \frac{\partial\delta\eta_i}{\partial\Delta\chi}\Delta\chi + \mathcal{O}(\Delta^2).
\end{aligned} \tag{6.2}$$

Using the high statistics TRKSIM samples and by dividing, one at a time, each of the volume sizes into 1/2 of the original size the derivatives in Eq. 6.2 are numerically evaluated. The correction evaluated in this way has been applied in Figures 6.1–6.3. The circles are the uncorrected points and the triangles are corrected.

This method has a tendency to overestimate the correction. This happens because the convergence, as the volume size goes to zero, is not exactly linear; the higher order terms in the expansion are not negligible. Also note that the derivatives are evaluated at a non-zero volume size. In Figures 6.1 – 6.3 there is also a set of points, shown as boxes, that correspond to a correction assuming purely quadratic convergence, i.e., there is no linear term. This correction is 2/3 of the correction with linear convergence in the volume size [1]. This shows that the systematic effects in Figures 6.1–6.3 are explained by the finite volume sizes.

[1] Using the known $f(x_1)$ and $f(x_1/2)$ and an assumed form of $f(x)$ $f(0)$ can be estimated. ($f(x_1)$ is known from the standard fit and $f(x_1/2)$ is known from the fit with the half volume.) If the convergence is linear, $f(x) = a + bx$, it follows by solving the linear

The next step is to correct the fit of the data. The first method is to simply use Figures 6.1–6.3 to get a correction. This has been done by sketching the function of the fitted values, R_1 and R_2 in Figure 6.1– 6.3, as a function of the generated values. By reading off the generated value that corresponds to the result of the fit of the data a correction is obtained. This corresponds to inverting the function. (The fit of the data gives $R_1 = 1.25 \pm 0.30$, $R_2 = 0.67 \pm 0.22$, and $\rho^2_{A_1} = 0.92 \pm 0.15$). The estimated corrections are summarized in Table 6.1.

As another approach, a sample of events has been drawn, according to the distribution predicted by the fit of the data, from the Monte Carlo sample used in the form factor fit. This sample of approximately 30,000 events is then fit and the effects of the finite volume sizes are obtained as the difference between the fitted parameters and the parameters the sample was drawn with. The procedure of drawing a subsample from the MC sample according to some given distribution is not an optimal procedure, since this gives a statistical correlation between the two samples.

The result of these two methods are summarized in Table 6.1. The corrections to be applied on the data are taken from the CLEOG fit and the statistical error is taken as the systematic uncertainty.

Method	δ_{R_1}	δ_{R_2}	δ_{ρ^2}
TRKSIM	-0.14	0.10	-0.01
CLEOG	-0.14 ± 0.048	0.07 ± 0.036	-0.03 ± 0.023

Table 6.1: The size of the volume size correction from the two methods based on either TRKSIM or CLEOG. The corrections are evaluated, in both methods, for the values of the form factors that are obtained from the fit of the data. The agreement between the two methods is good.

equation system that the correction is

$$\Delta_{\text{linear}} \equiv f(x_1) - a = 2(f(x_1) - f(x_1/2)). \tag{6.3}$$

Similarly, if the convergence is quadratic, $f(x) = a' + b'x^2$, the correction is

$$\Delta_{\text{quad.}} \equiv f(x_1) - a' = \frac{4}{3}(f(x_1) - f(x_1/2)), \tag{6.4}$$

or that $\Delta_{\text{quad}} = \frac{2}{3}\Delta_{\text{linear}}$.

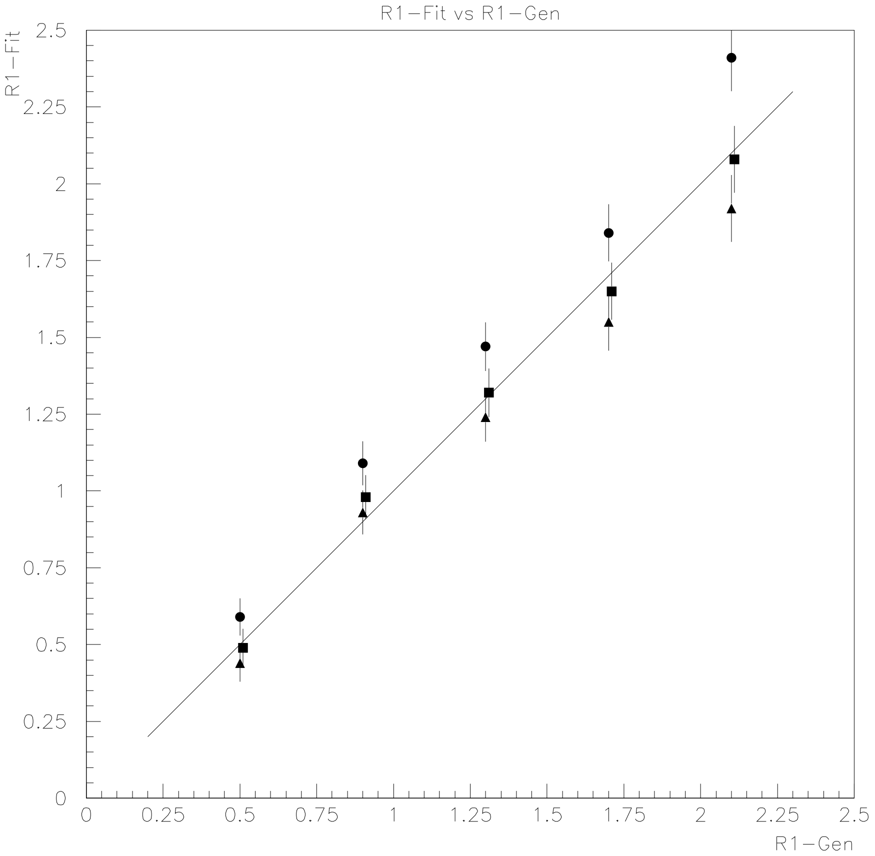

Figure 6.1: The fitted versus generated R_1 with and without the volume size correction. The uncorrected points are the circles and the corrected points are the triangles. The straight line is the line of $R_1^{\mathrm{fit}} = R_1^{\mathrm{gen}}$. This plot, based on fast Monte Carlo, is not the basis for the volume size correction. It is meant to demonstrate that the difference between the generated and fitted values are due to the finite volumes. The boxes are 2/3 of the correction that is applied for the triangles. This smaller correction corresponds to a convergence, as the volume size goes to zero, that is purely quadratic, that is, there is no linear term. The other parameters, R_2 and $\rho^2_{A_1}$, were fixed to the values obtained in the fit of the data ($R_2 = 0.65$ and $\rho^2_{A_1} = 1.0$).

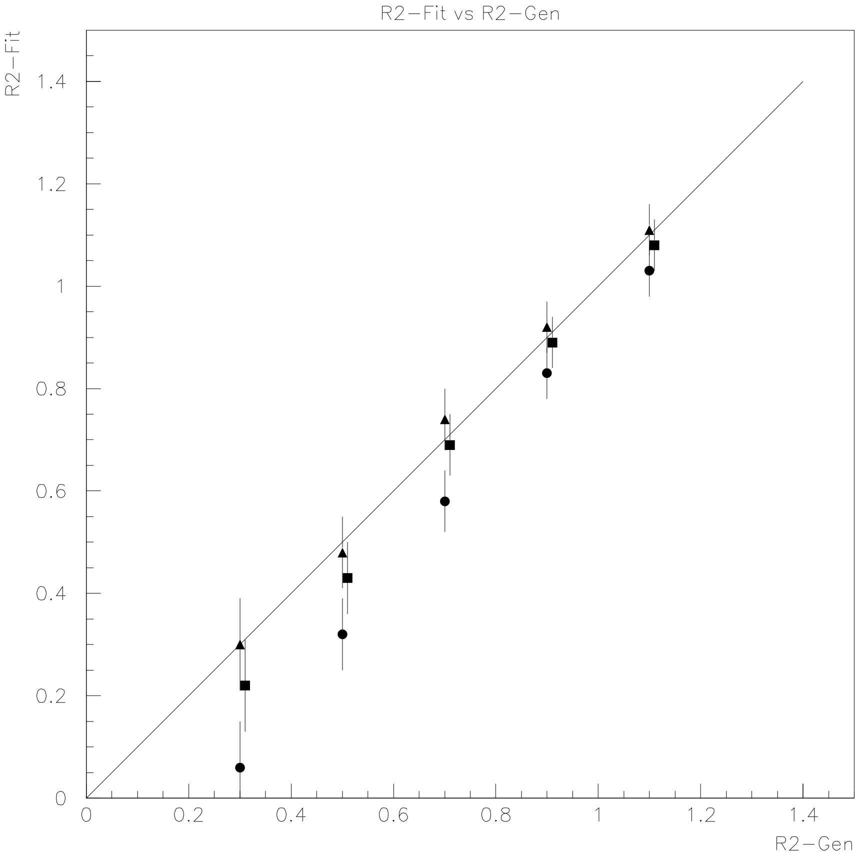

Figure 6.2: The fitted versus generated R_2 with and without the volume size correction. The uncorrected points are the circles and the corrected points are the triangles. The straight line is the line of $R_2^{\text{fit}} = R_2^{\text{gen}}$. This plot, based on fast Monte Carlo, is not the basis for the volume size correction. It is meant to demonstrate that the difference between the generated and fitted values are due to the finite volumes. The boxes are 2/3 of the correction that is applied for the triangles. This smaller correction corresponds to a convergence, as the volume size goes to zero, that is purely quadratic, that is, there is no linear term. The other parameters, R_1 and $\rho^2_{A_1}$, were fixed to the values obtained in the fit of the data ($R_1 = 1.3$ and $\rho^2_{A_1} = 1.0$).

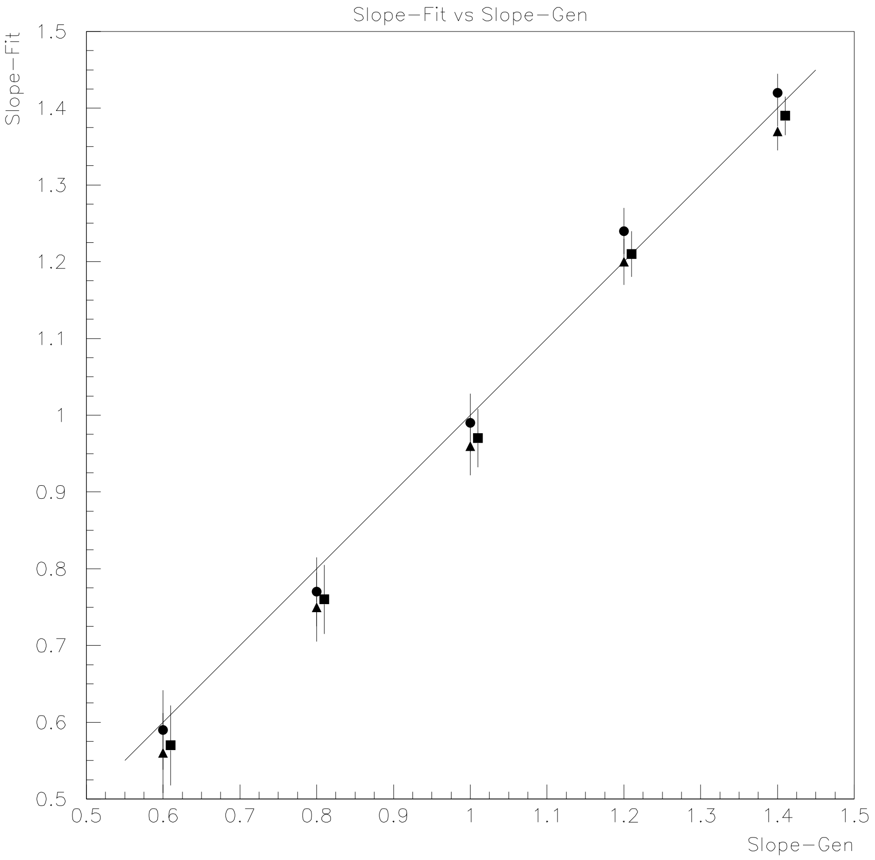

Figure 6.3: The fitted versus generated $\rho^2_{A_1}$ with and without the volume size correction. The uncorrected points are the circles and the corrected points are the triangles. The straight line is the line of $(\rho^2_{A_1})^{\text{fit}} = (\rho^2_{A_1})^{\text{gen}}$. This plot, based on fast Monte Carlo, is not the basis for the volume size correction. It is meant to demonstrate that the difference between the generated and fitted values are due to the finite volumes. The boxes are 2/3 of the correction that is applied for the triangles. This smaller correction corresponds to a convergence, as the volume size goes to zero, that is purely quadratic, that is, there is no linear term. The other parameters, R_1 and R_2, were fixed to the values obtained in the fit of the data ($R_2 = 0.65$ and $R_1 = 1.3$).

6.1.2 Monte Carlo statistics

Since the Monte Carlo sample used in the form factor fits is finite there will be a statistical error associated with the finite statistics. The number of events used in the Monte Carlo sample for the form factor fits is approximately 170×10^3 events, ≈ 200 times the size of the signal. On general grounds the statistical error on a quantity is proportional to $1/\sqrt{N}$, where N is the number of events. Therefore, the expected statistical error from the finite Monte Carlo sample is approximately $1/\sqrt{200}$ times the statistical error from the fit; giving $\sigma_{R_1} = 0.02, \sigma_{R_2} = 0.015, \sigma_{\rho^2} = 0.01$.

However, the fitting method allows for direct evaluation of the statistical errors from the finite Monte Carlo. By dividing the Monte Carlo sample into 4 subsamples and performing form factor fits using these independent samples the statistical uncertainty is evaluated by measuring the spread, the r.m.s., of the fit results. The estimated uncertainties are 0.03, 0.01, and 0.01 for R_1, R_2, and $\rho^2_{A_1}$ respectively. This is in agreement with the estimate of scaling the statistical error on the signal.

6.2 Backgrounds

The systematic uncertainties related to modeling the backgrounds are discussed below. First the systematic effects of the fake D^*s are discussed. This considers systematics related to using the δm sideband to model this background. Then the systematic effects related to the $B \to D^{**}\ell\nu$ background modeling are discussed. Last, the uncorrelated background is discussed.

6.2.1 Fake D^* background

The systematic error associated with using the δm sideband to model the fake D^* background breaks down to several different components. First the statistical error on the amount of the background is studied. Then the statistical error on the shape is also considered. Here a potential shape difference between the upper and lower half of the δm sideband is discussed. The sensitivity to the shape of the background is studied and a systematic error is assigned based on this sensitivity. Lastly, the possibility that the δm sideband might be underestimating the combinatorial background is considered, since events with a real D^* but a misreconstructed D^0 have a tendency to peak in δm. Each of these sources of systematic uncertainties are discussed in detail in the sections below.

The four different sources of systematic error that are considered in the modeling of the combinatorial background are added in quadrature to obtain the total contribution to the systematic error from modeling the combinatorial background. These contributions are summarized in Table 6.2. The statistical error in the shape is the dominating source of systematic uncertainty for the combinatorial background. The estimated uncertainties are 0.05, 0.03, and 0.02 for R_1, R_2, and $\rho^2_{A_1}$ respectively.

Contribution	R_1	R_2	$\rho^2_{A_1}$
Statistical uncertainty in rate	0.01	0.01	0.01
Statistical uncertainty in shape	0.04	0.02	0.01
Sensitivity to shape variations	0.01	0.01	0.00
True D^*s with fake Ds	0.02	0.02	0.01
Total	0.05	0.03	0.02

Table 6.2: The different contributions to the systematic error associated with the combinatorial background. The total error is obtained by adding the different contributions in quadrature.

6.2.1.1 Statistical error on the amount of combinatorial background

This error is straight forward to estimate. The amount of the combinatorial background is varied by $\pm1\sigma$. Note that the statistical errors on the amount of background in the 4 channels ($K\pi e$, $K\pi\mu$, $K\pi\pi^0 e$, and $K\pi\pi^0\mu$) are uncorrelated; therefore each submodes are varied $\pm1\sigma$ at the time according to Table 4.2 and the changes are added in quadrature. The systematic uncertainties assigned on the parameters in the form factor fit are 0.01, 0.01, and 0.01 for R_1, R_2, and $\rho^2_{A_1}$ respectively.

6.2.1.2 The statistical uncertainty on the shape of the combinatorial background

This error is evaluated in the same way as the statistical error on the Monte Carlo in Section 6.1.2. That is, the sideband sample from TRKSIM Monte Carlo is divided into 4 independent subsamples. Form factor fits are then performed using each of these samples to model the combinatorial background. The systematic uncertainty of 0.04, 0.02, and 0.01 is assigned on R_1, R_2, and $\rho^2_{A_1}$ respectively based on the spread (r.m.s.) of the fits.

The possibility that there is a shape difference in the kinematic variables away from the signal region has also been considered. To study this, fits have been performed using separately the upper and the lower half of the side band. The difference between the fit of the upper and the lower sideband is well within what is expected from the statistical uncertainty on the size of the sideband. Therefore, it is concluded that there is no significant shape difference between the upper and the lower sideband and no systematic uncertainty is assigned.

6.2.1.3 Sensitivity to the distributions in the δm sideband

To study the sensitivity to the shape of distributions taken from the δm sideband, the re-weighting technique is used on the events in the sideband. The re-weighting is such that it corresponds to a change of the shape by a 10% increase at the low end of the distribution for each variable and a 10% decrease at the other extreme. This change is applied to each of the variables, one at the time. In Table 6.3 the effect of this modification of the shape is shown. The sensitivity to this re-weighting is small compared to the statistical error. The table also shows the explicit form of the re-weighting factor. The 10% variation is motivated by fitting the data sideband for the tilt needed to obtain the best agreement with the TRKSIM sideband. It is found that for q^2, $\cos\theta_\ell$, and $\cos\theta_V$ this tilt is consistent with 0 with errors of about 10%. For χ the result of the fit is 0.23 ± 0.08. However, the sensitivity of the form factors with respect to tilts in χ is marginal. A 10% variation in χ is still applied. The estimated systematic uncertainties related to the uncertainty in the shape of the combinatorial background is 0.01, 0.01, and 0.00 for R_1, R_2, and $\rho^2_{A_1}$ respectively.

6.2.1.4 Fake D^0s in true D^*s

In Figure 6.4 the invariant $K\pi$ mass is shown for the signal region in δm (shown as the solid line) and the scaled sideband in δm (the dashed line). This issue discussed here is that the δm sideband might underestimate the fake D^* background since events with a true D^* and a fake D^0 have a tendency to peak in δm. What this means is that there is a $D^{*+} \to D^0\pi^+$ decay in the event and the correct slow pion is selected (the true D^*), but

Re-weighting factor	R_1	R_2	$\rho^2_{A_1}$
$1.0 + (q^2/q^2_{\rm max} - 0.5) \times 0.2$	$+0.011$	-0.009	0.000
$1.0 - (q^2/q^2_{\rm max} - 0.5) \times 0.2$	-0.012	$+0.010$	-0.001
$1.0 + \cos\theta_\ell \times 0.1$	$+0.004$	-0.003	0.000
$1.0 - \cos\theta_\ell \times 0.1$	-0.005	$+0.004$	-0.001
$1.0 + \cos\theta_V \times 0.1$	-0.005	$+0.005$	0.000
$1.0 - \cos\theta_V \times 0.1$	$+0.005$	-0.004	0.000
$1.0 + (\chi - 1.57) \times 0.064$	-0.002	$+0.002$	0.000
$1.0 - (\chi - 1.57) \times 0.064$	$+0.002$	-0.001	0.000

Table 6.3: The result of varying the shape of the δm distribution with the re-weighting factors quoted in the first column. The entries in the table quotes the difference between the fits with re-weighting and a standard fit. With this modification of the shape with a 10% change at the extreme values of the fit variables a small effect on the estimated form factors is obtained. The largest effect is in q^2. Summing these changes in quadrature gives the following contributions to the systematic error: $\sigma_{R_1} = 0.014$, $\sigma_{R_2} = 0.012$, and $\sigma_{\rho^2} = 0.001$.

the D^0 is misreconstructed (fake D^0). As seen in Figure 6.4 the signal region in δm shows a slight excess in the M_{D^0} sideband of about 16%.

A study is performed to show that the excess of events in the M_{D^0} sideband is from events that have a $D^{*+} \to D^0\pi^+$ decay in which the slow pion is correctly identified but the D^0 is misreconstructed. As is shown below, the D^0 is partially correctly reconstructed in the class of events that gives rise to the excess in the M_{D^0} sideband.

Using generic $B\bar{B}$ Monte Carlo events it is found that there is a clear excess of events in the δm signal region as compared to the sideband. This effect comes from signal events where the D^0 decayed either semileptonically or to K^+K^-. This study is done in a low M_{D^0} sideband; in the high M_{D^0} sideband there is an excess of $D \to \pi^+\pi^-$ events. The class of events that shows this excess has either 2 tracks coming from the D^0 with one track misidentified, e.g., a K^- is taken as a π^- in the Cabibbo supressed D decay, or it has 1 track from a secondary particle. In the Monte Carlo sample that was looked at, 317 events were scanned [37], see Figure 6.5, and in the signal region in δm there were 21 events of the type described above. In the sideband region there were only 3 events as compared to the expected $86 = 21 \times 4.1$, from scaling the signal region, 4.1 is the scaling factor between the sideband size and the signal region used in this study. The excess in this class of events is of the size needed to explain the excess in the δm signal region as compared to the scaled sideband.

The study described above was performed using a low $K\pi$ mass sideband. Studying the events in this class it is seen that the dominant type is $D^0 \to K^+K^-$. With one kaon misidentified as a pion, these events have a mass lower than 1.79 GeV/c^2. This has the consequence that these events do not contribute to an excess in the M_{D^0} signal region. Further, the plot of the data, Figure 6.4, shows that the excess appears to be most pronounced in the region below 1.8 GeV/c^2. The second largest class of events, semileptonic D decays, also seems to prefer a lower mass but there is not a distinct end of the mass distribution as in the case of the $D^0 \to K^+K^-$. Since the apparent excess of 16% in the δm sideband is largely from $D \to KK$ (60%) and the other main contribution, the semileptonic D decays, also indicates that there is more events below the signal region in M_{D^0}, no attempt will be made to try to account for this background in the fit, by, say, increasing the amount of combinatorial background. Further, it has also been verified for these events that the kinematic variables are reconstructed with a resolution comparable to the correctly reconstructed signal events. So if the acceptance for these events was the same as for the signal events they would be identical to signal events.

A systematic error is assigned by varying the amount of the combinatorial background by 16%. This gives the estimated systematic errors on each of the measured quantities as $\sigma_{R_1} = 0.02$, $\sigma_{R_2} = 0.02$, and $\sigma_{\rho^2_{A_1}} = 0.01$.

6.2.2 The contamination from $B \to D^{**}\ell\nu$

In Section 4.3.4 the background from $B \to D^{**}\ell\nu$, $D^{**} \to D^{*+}\pi$ were estimated. The effects of these backgrounds are estimated by adding the number of events for each of the backgrounds as estimated in Table 4.5. From these fits a correction is extracted which is applied to the form factor fits to correct for the neglected D^{**} background. These corrections are summarized in Table 6.4. The errors assigned are the errors propagated from the uncertainty on the normalization as estimated in Table 4.5.

Background mode	δR_1	δR_2	$\rho^2_{A_1}$
D_1	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.01
D_2^*	0.03 ± 0.04	-0.02 ± 0.04	0.01 ± 0.01
$^{1/2}P_1$	0.00 ± 0.00	-0.00 ± 0.00	0.01 ± 0.01
Non Res.	0.03 ± 0.07	-0.01 ± 0.02	0.00 ± 0.01
Total	0.07 ± 0.08	-0.03 ± 0.04	0.02 ± 0.02

Table 6.4: Corrections to obtain the right result for a fit that neglects the background from $B \to D^{**}\ell\bar{\nu}$. There are no significant corrections.

6.2.3 Uncorrelated background

The size of the uncorrelated background is fairly small, as is shown in Table 4.2. However, this background is, for kinematic reasons, peaked at low $\cos\theta_\ell$, which makes it have a very systematic impact on the estimation of R_1. The peaking at low $\cos\theta_\ell$ means that this background has an effect of making the forward–backward asymmetry too small or in other words it makes the fit underestimate R_1. This background is accounted for in the fit by adding a term to the probability distribution that accounts for this background. This shape is from generic $B\bar{B}$ Monte Carlo events, TRKSIM,

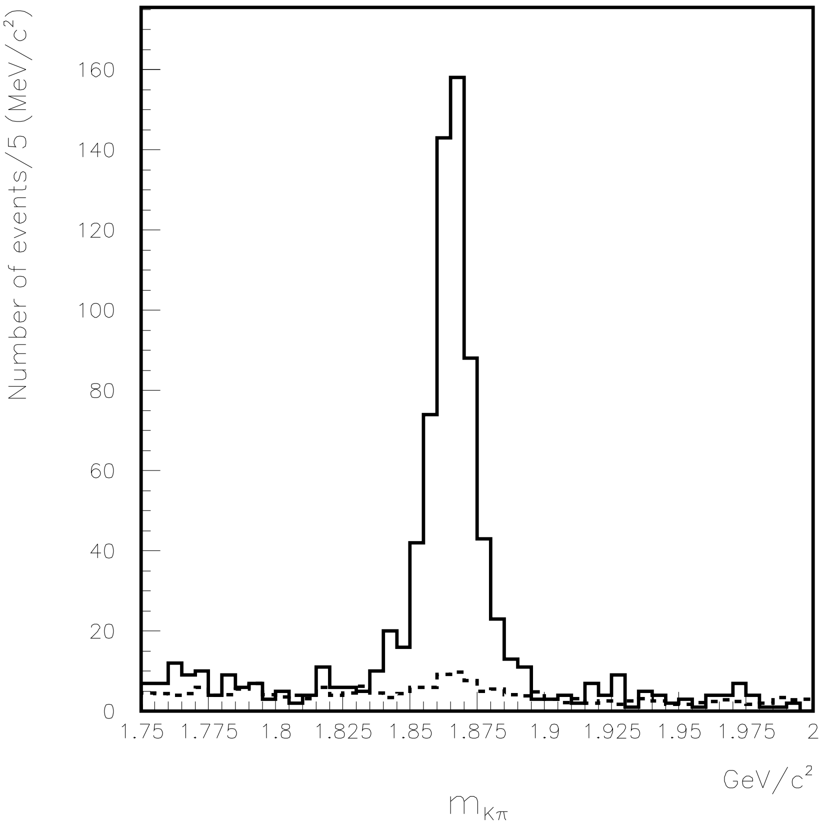

Figure 6.4: The invariant $K\pi$ mass for the signal region, solid line, and the δm sideband, dashed line. The scaled δm sideband appears to underestimate the background by approximately 16%. As discussed in the text this is largely an effect from misreconstructed D^0s which do not peak in the M_{D^0} signal region.

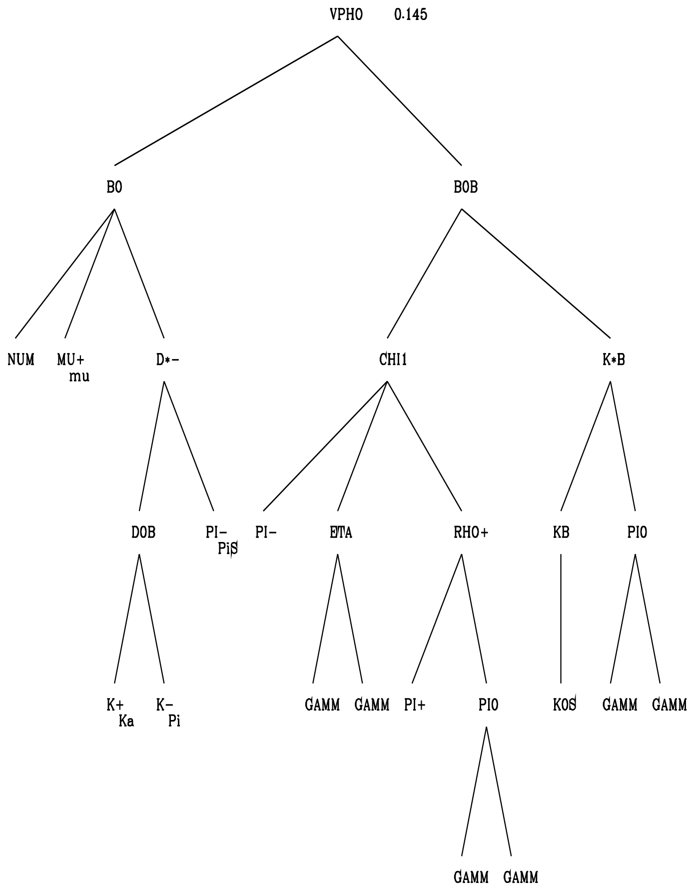

Figure 6.5: One of the 317 events in the low M_{D^0} sideband that was studied to understand the excess of events in the δm signal region. It was learned that this excess was from signal events with the lepton and the slow pion correctly reconstructed and the D^0 partially reconstructed, i.e., more than 1 track came directly from the D^0 or from a secondary particle decay. The dominant source was, in this low $M_{K\pi}$ sideband, the decay $D \to KK$; the event in the figure is of this type. The number to the right of "VPHO" is the reconstructed δm. The D^0 candidates had 1750 MeV/c^2 $< M_{D^0} <$ 1825 MeV/c^2

that, after passing all event selection criteria, are tagged[2] as uncorrelated. It is assumed that the distribution of these events are uncorrelated in the 4 kinematic variables. That this is a good approximation is motivated, primarily, by the fact that these events are distributed in a narrow region in $\cos\theta_\ell$ and that this distribution dominates the effect this background has on the fit of the data.

Adding this background into the fit changes R_1 by $+0.08$ and R_2 by -0.03. The effect on $\rho^2_{A_1}$ is negligible. The systematic error associated with the uncorrelated background is based on the statistical uncertainty on the amount of the uncorrelated background, as given in Table 4.2. Assuming that the statistical error in the $K\pi$ and the $K\pi\pi^0$ modes are uncorrelated, the following uncertainty in the form factors are found from a 1σ variation of the amount of the uncorrelated background $\sigma_{R_1} = 0.03$, $\sigma_{R_2} = 0.01$, and $\sigma_{\rho^2} = 0.01$. These variations are taken as the estimate of the of the systematic error associated with this background.

6.3 Monte Carlo modeling systematics

Below the systematic uncertainties related to the Monte Carlos modeling of the efficiency is discussed. In particular the uncertainty on the Monte Carlo modeling of the slow pion efficiency is discussed. There is also a discussion of the modeling of the cut on δm as a function of q^2.

6.3.1 Slow pion efficiency

The issue here is whether the Monte Carlo simulation is correct for the detection efficiency of the slow pion. Using data the efficiency has been measured [38]. The measurement of the efficiency has been carried out both in data and in Monte Carlo. The agreement between data and Monte Carlo is good, but exact tests are hard to do since the statistics is limited.

The slow pion efficiency is parametrized by the "Menary"-function as a function of the momentum. To test the sensitivity of the form factor fit with respect to the uncertainty in the slow pion efficiency a reweighting of the Monte Carlo according to the slow pion efficiency function is performed. The parameters of the Menary function is varied according to the 1σ uncertainties quoted in [38]. The resulting variation in the measured form factors is taken

[2]Tagged means that the Monte Carlo truth table was used to verify that the lepton and the D^* came from different Bs.

as the estimate of the systematic uncertainty. Table 6.5 summarizes the uncertainties from the slow pion efficiency on R_1, R_2, and $\rho^2_{A_1}$.

The largest uncertainty is, as might be expected, in the estimation of the form factor slope, $\rho^2_{A_1}$, since q^2 is strongly correlated with the slow pion momentum.

6.3.2 δm efficiency

Here the question is if the Monte Carlo is correctly modeling the width of the δm distribution. This might lead to a systematic effect if the Monte Carlo doesn't model the efficiency right for the cut on δm as a function, primarily, of q^2. At high q^2 the width of this distribution is larger since high q^2 means, in average, lower slow pion momentum and a worse momentum resolution. The cut on δm is approximately 2σ and is therefore fairly efficient. The uncertainty in the efficiency is studied using the re-weighting technique that has been applied in, e.g., the study of the slow pion efficiency. The change in the efficiency of the δm cut is calculated when changing the width of the δm distribution by the difference measured between data and Monte Carlo. The efficiency difference is used to re-weight the Monte Carlo. The width of the δm distribution is obtained by fitting the δm peak to a double gaussian in a low and a high q^2 bin. This leads to a small uncertainty in the estimated form factors, see Table 6.5.

6.4 Summary of systematic errors

In the previous sections several sources of systematic errors have been considered. To obtain the total contribution the individual components are simply added in quadrature. All sources of systematic errors that are considered are summarized in Table 6.5. This measurement is statistics dominated; the estimated systematic error is typically 1/3 of the statistical error. The estimated systematic errors are 0.12, 0.07, and 0.06 for R_1, R_2, and $\rho^2_{A_1}$ respectively. The largest contribution is from the uncertainty in the $B \to D^{**}\ell\nu$ background followed by the uncertainties related to the fitting method and the modeling of the fake D^* background.

	R_1	R_2	$\rho^2_{A_1}$
Finite volumes	0.05	0.04	0.03
MC statistics	0.03	0.01	0.01
Fake D^*s	0.05	0.03	0.02
$B \to D^{**}\ell\nu$	0.08	0.04	0.02
Uncorr. Bkgrnd.	0.03	0.01	0.01
Slow pion eff.	0.01	0.01	0.04
δm eff.	0.01	0.01	0.02
Lepton eff.	0.02	0.02	0.01
Sum	0.12	0.07	0.06

Table 6.5: The estimated systematic errors.

Chapter 7

Results and conclusions

7.1 Results of the form factor fits

Here the result of the form factor fits together with the estimated systematic errors and corrections are presented.

In Figure 7.1 the one dimensional projections of the fit are shown. The fit agrees well with the data in all four projections. In particular it is worth noticing that the assumed linear form of $h_{A_1}(w)$ does an excellent job of describing the q^2 distribution. In the χ-projection the statistics are sufficient to start seeing the $\cos 2\chi$ term.

Figure 7.2 shows the $\cos\theta_V$ and $\cos\theta_\ell$ distributions in a low and a high q^2 bin. For low q^2, the D^* is moving fast in the B rest frame, the D^* is longitudinally polarized. This gives rise to a $\cos^2\theta_V$ distribution at $q^2 = 0$, while for $q^2 = q^2_{\rm max}$ the $\cos\theta_V$ distribution is uniform. However, at $q^2_{\rm max}$ for $\cos\theta_V$ close to $+1$, the effects of acceptance due to the poor detection efficiency of the slow pion are severe. When $\cos\theta_V$ is close to $+1$ the D is moving forward and the π backward in the D^* rest frame.

A very interesting plot is shown in Figure 7.3. The χ distribution is shown in three bins of $\cos\theta_V$. For $\cos\theta_V < -0.5$ the χ distribution has more events at large χ than at low χ. While for $\cos\theta_V > 0.5$ the opposite pattern is observed. This is in agreement with the expectation from $(V-A)\times(V-A)$ at both W vertices. However, in a theory where one of the W-vertices were $(V+A)$ this pattern would be reversed.

Figures 7.4 and 7.5 show a series of 2-dimensional projections. Though somewhat hard to interpret, these plots indicate a good agreement between the data and the fit. The fit describes the data well, e.g., in the plot of $\cos\theta_V$ versus q^2, the $\cos^2\theta_V$ distribution at low q^2 is clear.

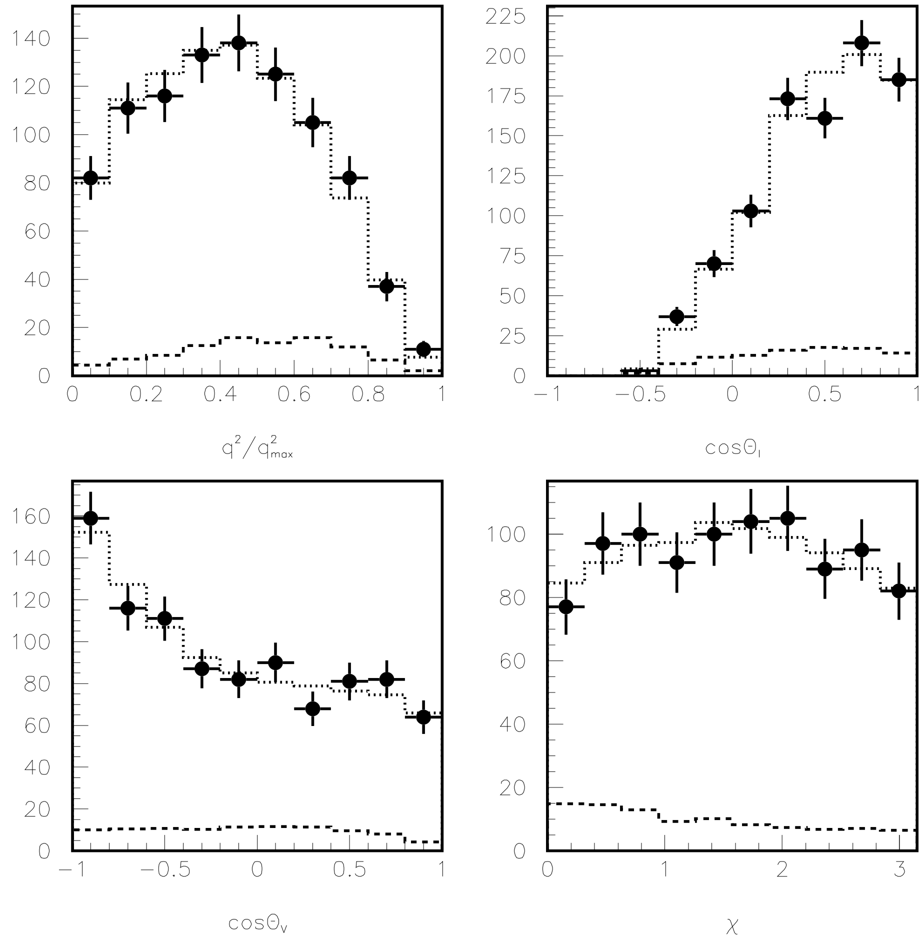

Figure 7.1: The one dimensional projections of the fit. The data are indicated by the points with error bars and the fit is indicated by the dashed line. The backgrounds from fake D^* and uncorrelated D^*-lepton pairs are indicated by the solid line.

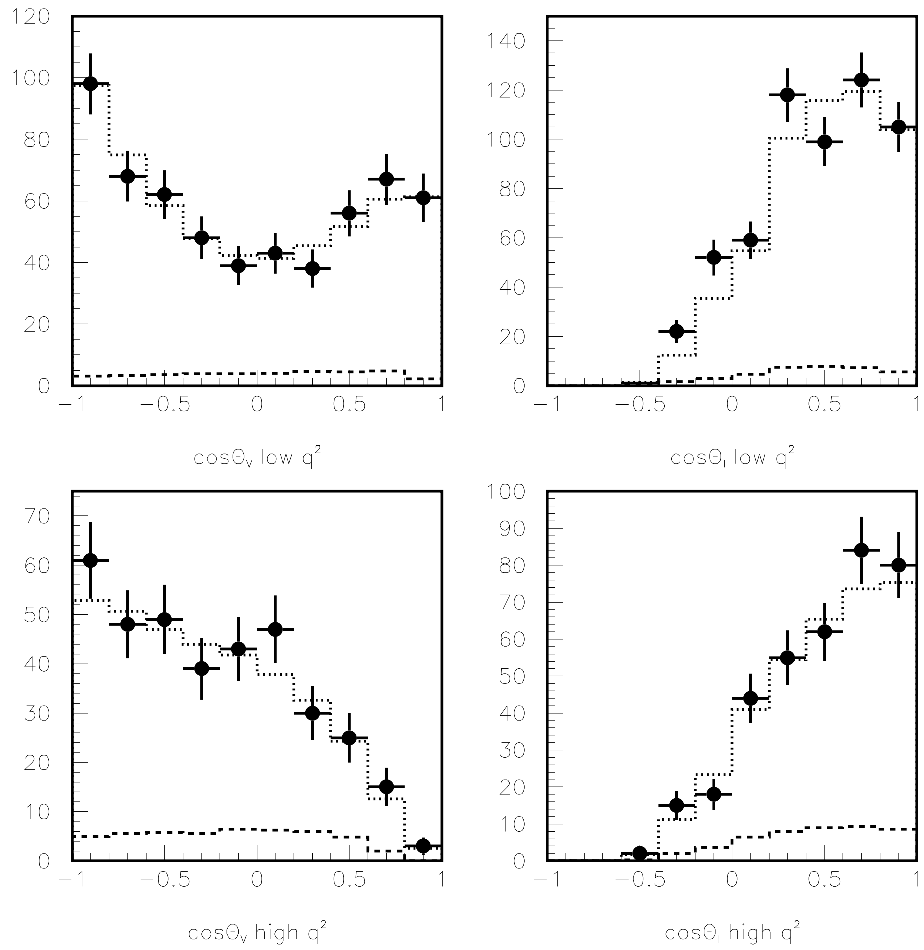

Figure 7.2: Distributions of $\cos\theta_V$ and $\cos\theta_\ell$ in low and high q^2 bins. The data are indicated by the points with error bars and the fit is indicated by the dashed line. The backgrounds from fake D^* and uncorrelated D^*-lepton pairs are indicated by the solid line. In the low q^2 bin, the $\cos^2\theta_V$ distribution that is expected at $q^2 = 0$ is seen.

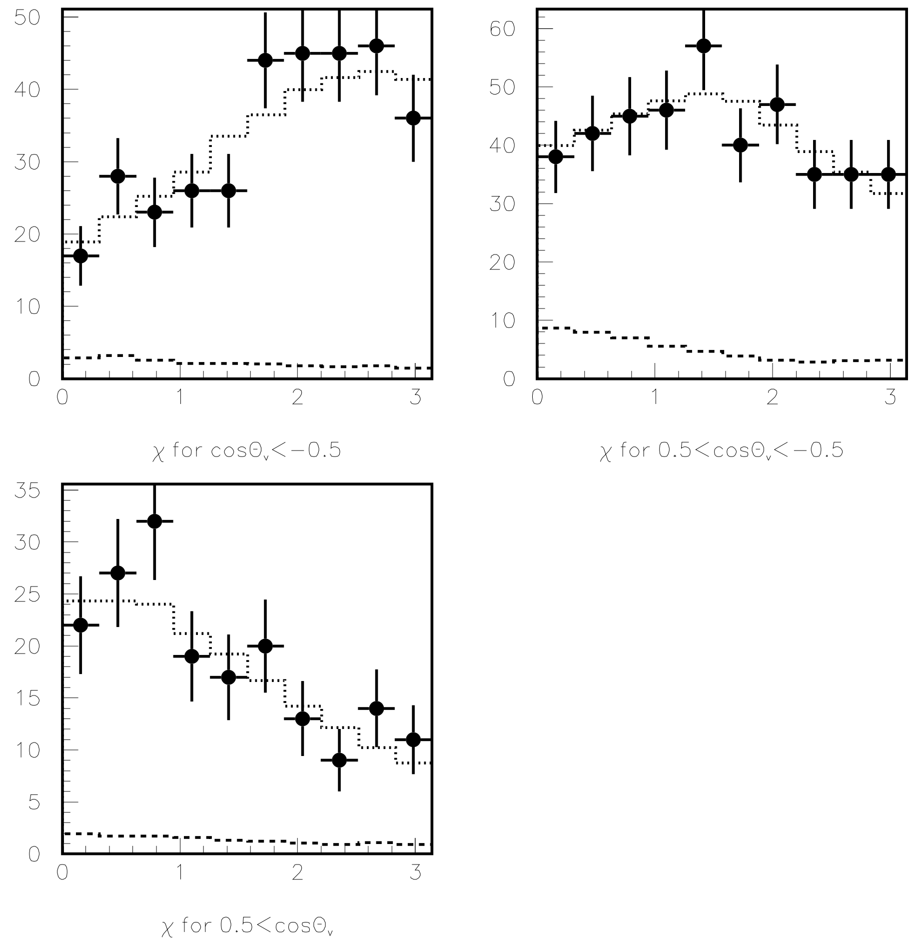

Figure 7.3: χ in three $\cos\theta_V$ bins. These plots are interesting since the distributions in the χ variable are sensitive to the form of the weak current. If the Wbc vertex had a $V+A$ coupling the shape distributions for $\cos\theta_V < 0$ and $\cos\theta_V > 0$ would be reversed, which clearly is in disagreement with what is observed in the data.

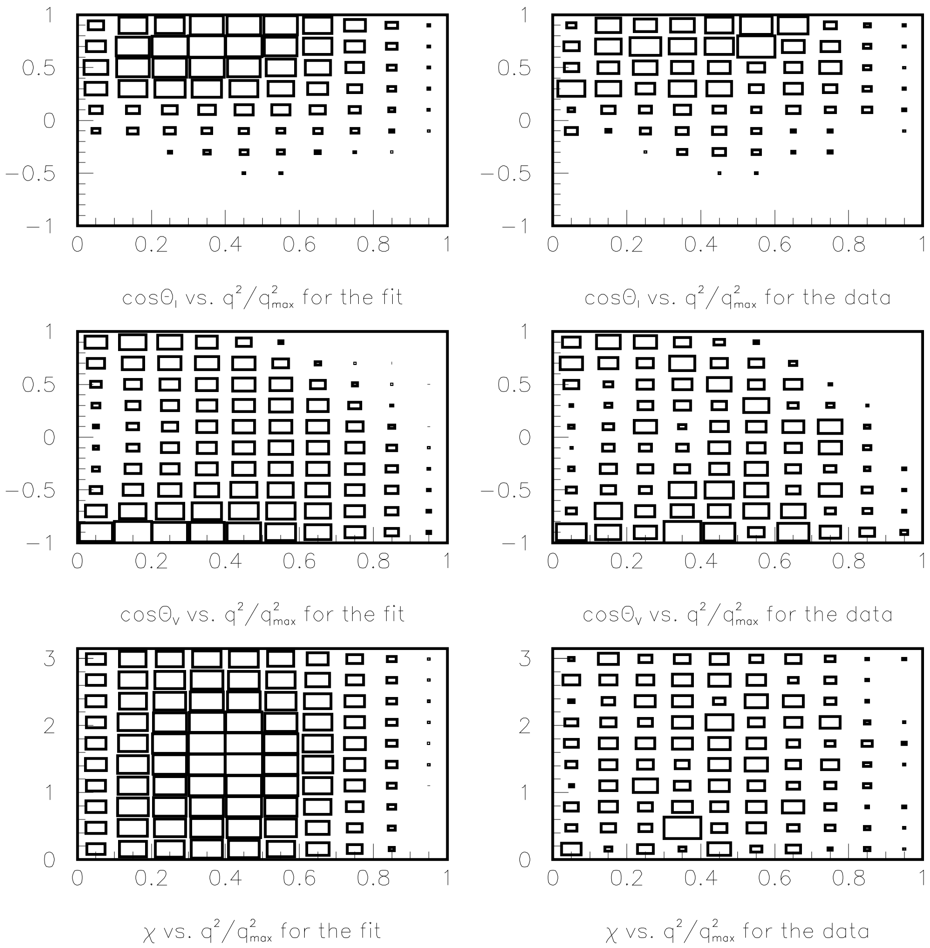

Figure 7.4: 2-dimensional projections of the fit. The fit is shown to the left and the data to the right. At low q^2 the $\cos^2 \theta_V$ distribution is clear.

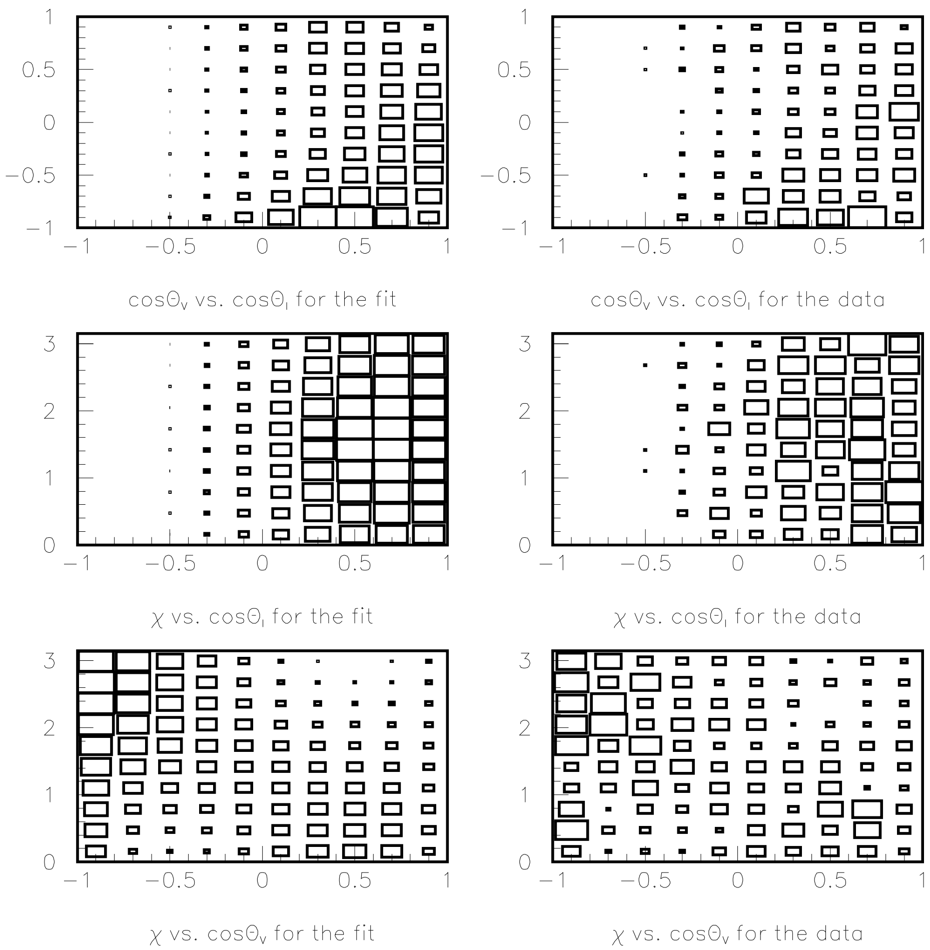

Figure 7.5: 2-dimensional projections of the fit. The fit is shown to the left and the data to the right. The structure in the plot of χ versus $\cos\theta_V$, an enhancement of events in the upper left and lower right corners, are an effect of $(V-A)\times(V-A)$ at both W vertices. If one of the vertices had a $(V+A)$ coupling the pattern would be reversed. This is clearly not in agreement with what is observed.

There are two corrections that are applied to the fits. The first is the effect of the finite search volumes and the second is the effect of the D^{**} background. Both of these corrections were described in great detail in the previous chapter.

Correction	δR_1	δR_2	$\delta \rho^2_{A_1}$
Finite volumes	-0.14	0.07	-0.03
$B \to D^{**}\ell\nu$	0.07	-0.03	0.02
Total	-0.07	0.04	-0.01

Table 7.1: The corrections that are applied to the fit. For a discussion of the origin of these corrections see Sections 6.1 and 6.2.2.

Applying these corrections, give the final result $R_1 = 1.18 \pm 0.30 \pm 0.12$, $R_2 = 0.71 \pm 0.22 \pm 0.07$, and $\rho^2_{A_1} = 0.91 \pm 0.15 \pm 0.06$.

7.1.1 Alternative fits

A series of alternative fits are performed in which the assumptions have been modified. The first two such alternative fits have R_1 and R_2 modified with the w dependence according to Neubert [24] and Close and Wambach [25]. The w dependence calculated in these references are the effects of heavy quark symmetry breaking and α_s corrections and is found to be relatively mild. In particular, the calculation by Close and Wambach finds a very small w dependence on the form factor ratios. Assuming the w dependence in these references, the small changes in R_1 and R_2 according to rows 2 and 3 in Table 7.2 are found. The first entry in the table is the standard fit with a linear Isgur-Wise function and constant form factor ratios. The value of the form factor slope, $\rho^2_{A_1}$, is unchanged with this alternative form of the form factor ratios.

The last three rows in the table contain the results of fits with alternative forms of the Isgur-Wise function. Three different forms that have been suggested in the literature have been tried. The first form is the exponential as used in the ISGW model [18]; the next form is suggested by the relativistic oscillator model [39]. Last, the pole form suggested by Neubert [24] is tried. On general grounds the Isgur-Wise function is expected to have a positive curvature. The Isgur-Wise function is expected to exhibit a pole at $w = -1$ (or $q^2 = (m_b + m_c)^2$), and fall off to zero as $w \to \infty$. All three alternative forms of the Isgur-Wise function considered here have a positive curvature and the last two forms, the relativistic oscillator and the pole form, exhibit the pole at $w = -1$. (Note that the $w = -1$ configuration is not kinematically accessible in semileptonic decays but is in scattering [24].) All forms considered have the Taylor expansion

$$h_{A_1}(w) = 1 - \rho^2_{A_1}(w-1) + \mathcal{O}((w-1)^2). \tag{7.1}$$

Thus, the extracted parameter $\rho^2_{A_1}$ is the slope of the Isgur-Wise function at $w = 1$. However, since the fit is sensitive to the average slope over the full w range, the obtained values for $\rho^2_{A_1}$ in the alternative fits are different. In particular they are larger than the linear form. As pointed out by Neubert [3] the fit using a linear form for the Isgur-Wise function gives a conservative lower limit of the slope at $w = 1$. The extracted form factor ratios remains virtually unchanged in the fits with alternative forms of the Isgur-Wise function.

$h_{A_1}(w)/h_{A_1}(1)$	$R_1(w)$	$R_2(w)$	R_1	R_2	$\rho^2_{A_1}$
$1-\rho^2_{A_1}(w-1)$	R_1	R_2	$1.18 \pm 0.30 \pm 0.12$	$0.71 \pm 0.22 \pm 0.07$	$0.91 \pm 0.15 \pm 0.06$
$1-\rho^2_{A_1}(w-1)$	$R_1^{\rm Neub}$	$R_2^{\rm Neub}$	$1.24 \pm 0.30 \pm 0.12$	$0.67 \pm 0.22 \pm 0.07$	$0.90 \pm 0.15 \pm 0.06$
$1-\rho^2_{A_1}(w-1)$	$R_1^{\rm CW}$	$R_2^{\rm CW}$	$1.20 \pm 0.30 \pm 0.12$	$0.70 \pm 0.22 \pm 0.07$	$0.91 \pm 0.15 \pm 0.06$
$\exp[-\rho^2_{A_1}(w-1)]$	R_1	R_2	$1.21 \pm 0.32 \pm 0.12$	$0.72 \pm 0.22 \pm 0.07$	$1.27 \pm 0.29 \pm 0.12$
$\frac{2}{w+1}\exp[(1-2\rho^2_{A_1})\frac{w-1}{w+1}]$	R_1	R_2	$1.21 \pm 0.32 \pm 0.12$	$0.71 \pm 0.22 \pm 0.07$	$1.53 \pm 0.36 \pm 0.14$
$(\frac{2}{w+1})^{2\rho^2_{A_1}}$	R_1	R_2	$1.21 \pm 0.32 \pm 0.12$	$0.70 \pm 0.22 \pm 0.07$	$1.42 \pm 0.32 \pm 0.13$
$R_1^{\rm Neub} = R_1[1-0.16(w-1)+0.07(w-1)^2]$			$R_2^{\rm Neub} = R_2[1+0.19(w-1)-0.05(w-1)^2]$		
$R_1^{\rm CW} = R_1[1-0.06(w-1)]$			$R_2^{\rm CW} = R_2[1+0.04(w-1)]$		

Table 7.2: Fits for the form factors with alternative forms of $h_{A_1}(w)$. The first entry in the table is the standard fit with a linear form of $h_{A_1}(w)$ and constant form factor ratios. The next two entries have w dependent form factor ratios according to theoretical predictions. The last three entries use alternative forms of $h_{A_1}(w)$; the extracted values of $\rho^2_{A_1}$ are sensitive to these forms. Note that the form factor ratios are not very sensitive to the form of $h_{A_1}(w)$.

7.1.2 The D^* polarization and $A_{\rm FB}$

Using the form factor slope and the form factor ratios obtained from the standard fit the forward-backward asymmetry, $\bar{A}_{\rm FB}$, and the D^* polarization,

$\bar{A}_{\rm pol}$ are calculated

$$\bar{A}_{\rm FB} = 0.197 \pm 0.033 \pm 0.016, \quad (7.2)$$
$$\bar{A}_{\rm pol} = 1.55 \pm 0.26 \pm 0.13. \quad (7.3)$$

$\bar{A}_{\rm FB}$ and $\bar{A}_{\rm pol}$ were defined in Section 2.5. The correlation coefficient between these measurements is found to be $\rho = 0.35$. These numbers can be compared to an ARGUS measurement [17] of $\bar{A}_{\rm FB} = 0.20 \pm 0.08 \pm 0.06$ and $\bar{A}_{\rm pol} = 1.1 \pm 0.4 \pm 0.2$. For the value of $\rho^2_{A_1} = 0.9$, Neubert calculates [24] $\bar{A}_{\rm FB} = 0.22$ and $\bar{A}_{\rm pol} = 1.40$.

7.2 Goodness of fit

A lot of effort has been put into trying to understand the goodness of the fit. However, goodness of fit has proven to be a very hard question to solve for multi-dimensional distributions. There are no standard methods, as there are for 1-dimensional distributions. One way that is often used to estimate the goodness of fit is to compare the measured $\log(\mathcal{L})$ with the distribution of the same quantity for a set of fits of Monte Carlo samples. This procedure is applied to the combined fit and it results in a confidence level of 14%, see Figure 7.6. The obtained confidence level of 14% is reasonable good. It should also be pointed out that the likelihood distribution is not necessarily a measure of the goodness of the fit. A counter example is found in the example of a likelihood fit to an exponential; where the $\log(\mathcal{L})$ is just the mean of the data points and does not provide a measure of the goodness of the fit.

7.3 Summary and conclusions

The form factors for the process $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}$ have been measured. The measurmant was based on 940 events with an estimated background contamination of 17%. The largest background, in both the $D^0 \to K^-\pi^+$ and the $D^0 \to K^-\pi^+\pi^0$ channel, were fake D^*. In the four dimensional fit the three parameters R_1, R_2, and $\rho^2_{A_1}$ have been determined. The fit describes the data well in all projections and a confidence level of 14% is estimated for the fit. The two form factor ratios $R_1 = 1.18 \pm 0.30 \pm 0.12$ and $R_2 = 0.71 \pm 0.22 \pm 0.07$ are in good agreement with the predicted corrections to the heavy quark symmetry limit; Neubert [24] predicts $R_1(1) = 1.35$ and

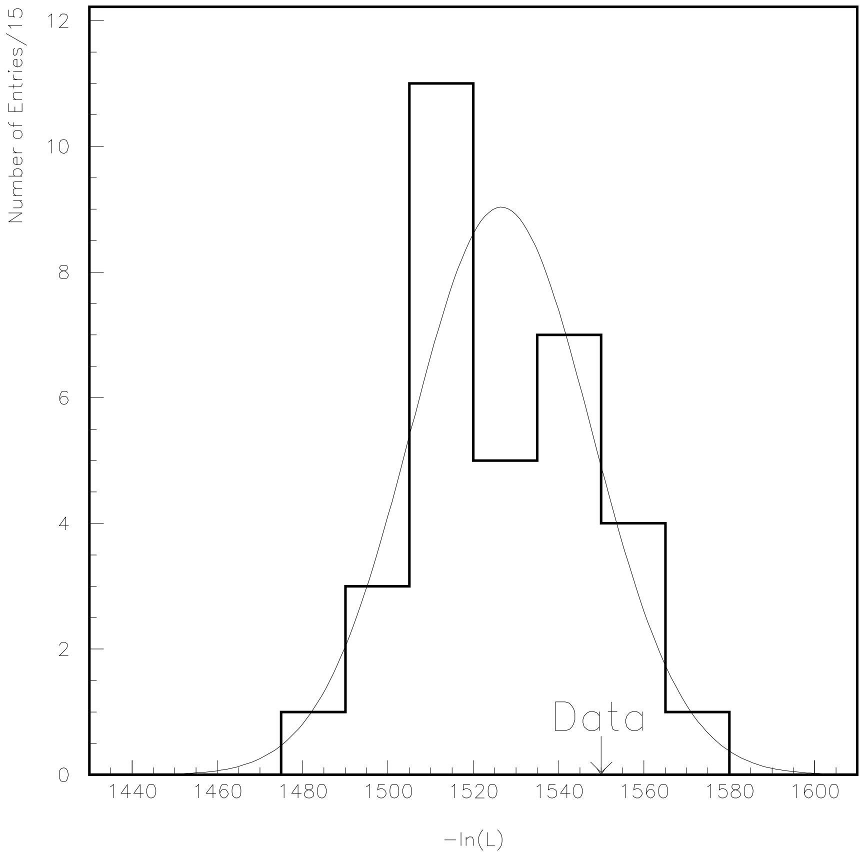

Figure 7.6: The distribution of $-\log(\mathcal{L})$ from Monte Carlo experiments and the fit of the data. The likelihood is defined as the fraction of the area of the distribution from the Monte Carlo that is higher (worse) then the fit of the data. The curve in the plot is the best fit of a gaussian to the Monte Carlo. Using this best fit the likelihood is evaluated to be 14%.

$R_2(1) = 0.79$, Close and Wambach [25] give $R_1(1) = 1.15$ and $R_2(1) = 0.92$. The results obtained are also consistent with the predictions of the ISGW2 model, $R_1 = 1.27$ and $R_2 = 1.01$.

The measured form factor slope $\rho^2_{A_1} = 0.91 \pm 0.15 \pm 0.06$ is in the range of predictions from lattice QCD and QCD sum rules [13]. The result is also consistent with the measurement of $\hat{\rho}^2 = 0.84 \pm 0.12 \pm 0.08$ from CLEO [6].

The largest systematic error is from the modeling of the $B \to D^{**}\ell\nu$ background. Even though this is not the largest background it has the largest uncertainty since it is not well understood. Other significant contributions of systematic errors are the modeling of the fake D^* background and the fitting method itself.

In the fit it was assumed, as discussed in Section 2.7, that R_1 and R_2 are constant and that $h_{A_1}(w)$ is linear in w. These assumptions were tested in Section 7.1.1 and it was found that R_1 and R_2 were not sensitive to either the form of $h_{A_1}(w)$ or a w dependance of R_1 and R_2 according to theoretical predictions. However, it was found that $\rho^2_{A_1}$ is sensitive to the assumed form of $h_{A_1}(w)$.

The fact that the breaking of the heavy quark symmetry is small and that R_1 and R_2 do indeed seem to be close to unity build confidence in using HQET as a starting point for measuring $|V_{cb}|$.

Future measurements, at any of the B-factories that will come on-line in a few years, will provide more accurate measurements. With 10-fold statistics the systematic uncertainty will be significant as compared to the statistical uncertainty. However, by then, the resonant components of the $B \to D^{**}\ell\nu$ decays should also be better understood. An asymmetric B-factory might also be able to reduce the background contamination by using the separate B-vertices to reduce the combinatorial background.

Part II

Billoir fit

Chapter 8

Introduction

The CLEO II detector has 3 tracking devices, the PT, the VD, and the DR, as discussed in Section 3.4. These devices measure the positions along the trajectory of a particle. For the wire chambers these position measurements are the distance from the track to the wire, the drift distance. For the cathods the position measurements are z positions measured at the inner and outer radius of the DR and the VD. When the detector is readout after an interaction all channels in the detector that detected a particle, i.e., passed a treshold, are read out.

The task of track reconstruction is twofold; the first task is track finding or pattern recognition, the second is track fitting. The goal of pattern recognition is to group the hits coming from the same track together. It should also solve ambiguities, e.g., the left-right ambiguity for the drift distance. An example of an event is shown in Figure 8.1. By eye, it is very simple to identify the tracks in the event. However, the task of writing computer programs to find the tracks, correctly and efficiently, is hard. Most algorithms are based on trying a lot of combinations. The algorithm used in the CLEO II reconstruction code looks for track segments in superlayers. A superlayer is several layers grouped together. These segments are then linked together to form the complete tracks. The task of pattern recognition is complicated by the presence of noise and inefficiencies in the detector. Another problem is curlers, tracks with a radius of curvature so small that they don't leave the detector untill they reach the end of the detector along the z-axis. The Billoir fitter described in this dissertation can be used as a tool for patter recognition. However, in the work presented here it is only used as a track fitter.

Track fitting is the second step in track reconstruction. Here the hits that were found in the pattern recognition are fitted to obtain optimal estimates of the momentum and the production point. This also includes estimating the errors on the track parameters that describes the trajectory. For the majority of the tracks the track parameters of interest are at the origin, the interaction point, where the particles were produced. However, there are particles that comes from secondary vertices. They are decay products of a long lived particle, e.g., a K_s or a Λ.

To obtain the optimal estimate of the track parameters the fitter has to account for the effects of multiple scattering and energy loss that the particle was subject to as it propagated outward through the detector. This becomes a fairly time consuming operation and this is the main reason that in the pattern recognition step simple, fast, fitters are used for the matching of the track segments.

The old track fitter for CLEO II performed a simple χ^2 fit to the hits found by the pattern recognition; it ignored the effects of energy loss and multiple scattering. The effects of multiple scattering on the variance matrix are incorporated after the fit is performed [40]. The effects of energy loss are corrected by the user of the track by a package of energy loss correction routines [41]. This correction is based on the transverse momentum of the track (how far out in the detector the track reaches). The correction has an ugly property of being discontinous at a transverse momentum of about 65 MeV/c[1].

CLEO has recently installed a new silicon vertex detector and data taking has begun. This vertex detector will provide very precise position measurments close to the interaction point. To make full use of the data taken with this vertex detector it is important that the track fitter can handle the effects of multiple scattering and energy loss correctly since these effects will be important even for tracks with a medium to high momentum. Therefore, the fitter was designed to be used in an environment with a silicon vertex detector as well as the PT, VD, and DR configuration.

The Billoir fitter described here provides the tool to perform track fits that account for the effects of energy loss and multiple scattering. This leads to a better understanding of the error matrix and better central values since

[1]For tracks with a transverse momentum below 65 MeV/c the momentum is assumed to be determined by the VD while for tracks with a transverse momentum above 65 MeV/c the momentum is assumed to be determined by the DR. Depending on where the momentum is assumed to be determined there are different energy loss corrections depending on how much material the track passed through before it reaches the detector element that determined the momentum.

it performs an accurate treatment of the energy loss.

8.1 Goals for the Billoir fitter

The goal of this project is to write a final track fitter that will correctly treat the effects of multiple scattering and energy loss. This also means that it will have to do separate fits for the five different particle hypothesis: e, μ, π, K, and p, as their interactions with matter are different. Further, there is a time constraint for a fit. The old final fitter, `TF3FIT`, took approximately 15 ms per fit, on a DEC Alpha 600. A goal was set to not take significantly longer time with the new fitter. However, the new fitter has to perform five fits per track, one fit for each particle hypothesis, and account for the material, so the code has to be efficient in order to meet this goal. Some effort has been put into code optimization. Simple things, like to avoid looping over indices, save a significant amount of time.

Another goal is that the fitter should use the material description in `CLEOG`, the `GEANT` [36] based detector simulation package.

Figure 8.2 illustrates the different building blocks that make up the Billoir fitter. At the heart of the method is the operation of adding a hit to the track. This operation needs information about the hits and the track parameters. The second operation that is essential for the fitter is the transport of the track parameters. The operation of transport takes track parameters at a given point and expresses them at another point. This requires knowledge about the B-field, the material, particle hypothesis, and so on.

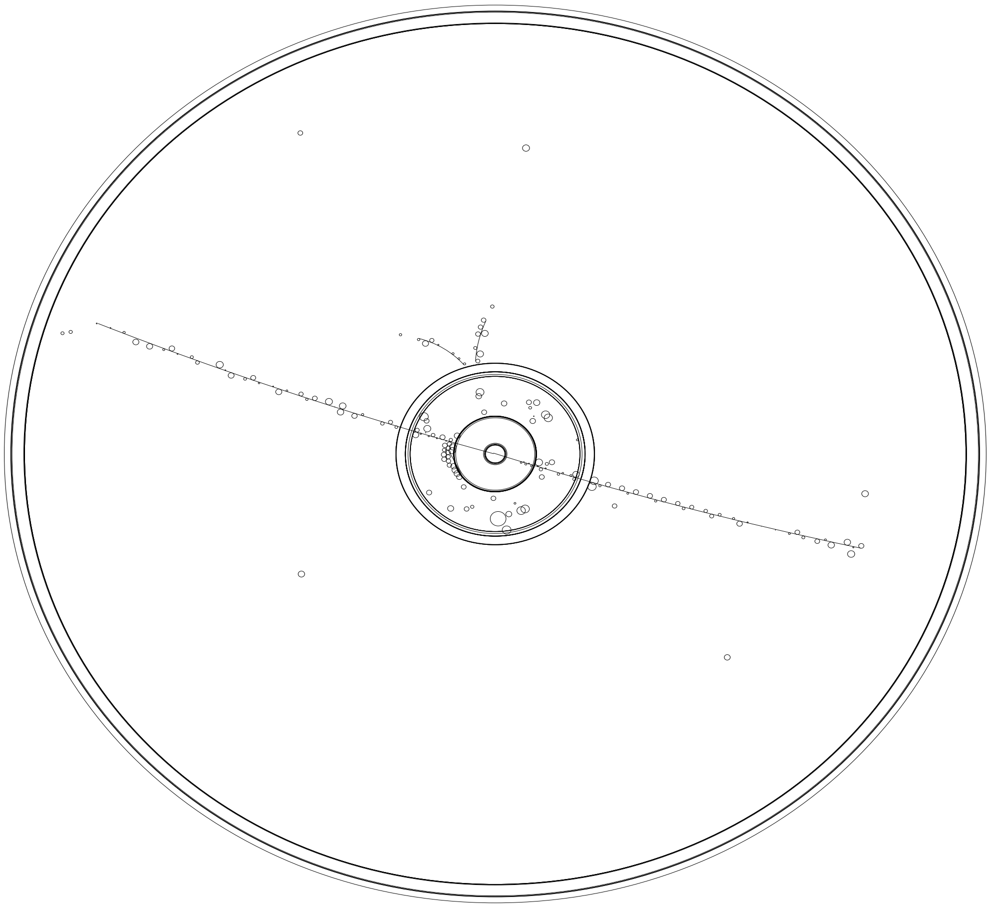

Figure 8.1: An example of en event collected by the CLEO II detector. Only hits on axial wires, wires parallel to the z-axis are shown in this display. This event shows a radiative Bhabha or mu-pair in which the photon converts in the material between the VD and the DR.

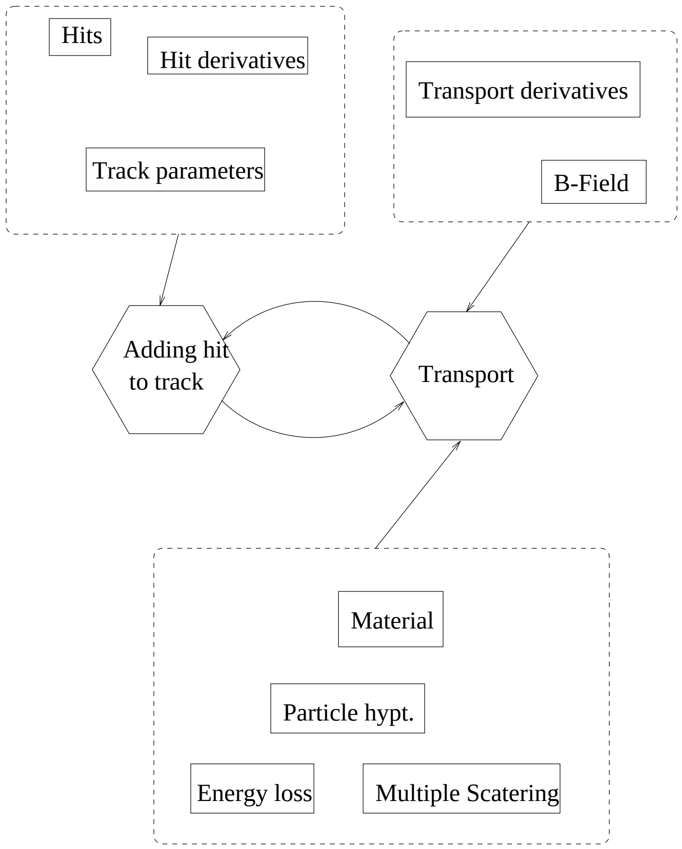

Figure 8.2: The task of estimating the optimal track parameters and simultaneously correctly estimate the variance matrix for the estimated track parameters involves a solid understanding of both the detector: resolutions, alignment, magnetic field, etc. and the interactions the produced particles have with the material in the detector: multiple scattering and energy loss.

Chapter 9

The Billoir fit

9.1 Motivation for the Billoir fitter

This section describes how the Billoir fit [42, 43], also known as a Kalman filter [44], works. The idea applied to a track fit is quit simple and beautiful. First, the idea of the Billoir fit is introduced in Section 9.2. A more precise, mathematical, formulation is given in Section 9.3. In the discussion in these two sections the issue about starting the fit has been avoided. In this implementation, the approach of linearizing the fit around a seed track is taken to obtain stability at the start of the fit. Exactly what this means is described in Section 9.4.

Before discussing how the Billoir fitter works the advantages of the method is discussed. This discussion will be on the form of a comparison with a standard χ^2-fit for the extraction of track parameters from a set of position measurements along the particle trajectory.

When a charged particle goes through a uniform magnetic field it traces out a helix in space. (A uniform field is considered for simplicity but it is not necessary for the discussion.) The particle's trajectory is a function of some set of initial parameters, e.g., its momentum and position when produced. The task of track fitting, in general, is to determine these initial parameters from a set of position measurements along the track. When a particle does not interact with material, an optimal estimate of the initial parameters is given by a simple χ^2-fit to the measured positions. It is assumed that measurements are normally distributed, so that the estimator is optimal (and unbiased). However, if there is material along the track, the (charged) particle will interact, for example by multiple scattering. The simple χ^2-fit must then be modified in order to handle this. A more general version of the χ^2-fit which can handle this requires an inversion of a (symmetric) $N \times N$ matrix, where N is the number of measurements along the track. This becomes a time consuming operation as the computation time needed to invert an $N \times N$ matrix is, for practical purposes, $\mathcal{O}(N^3)$. This is the main disadvantage with the general χ^2 method. The other disadvantage is that it is difficult to calculate the required derivatives of the measurements with respect to the track parameters. This is especially so when the trajectory has a complicated form, for example when there is energy loss along the track or if the magnetic field is not uniform. This problem is solved in the Billoir fitter by transporting the track to the hits.

The Billoir fitter described here is intended for use in a uniform solenoidal field, but is modified to handle smaller non-uniformities in the magnetic field. As will be shown below, the Billoir fitter provides an algorithm for doing the fits with a time complexity of $\mathcal{O}(N)$.

9.2 The idea behind the Billoir fit

From a set of measurements, in this case positions along a track, a set of parameters, η, needs to be determined. The idea is to add the information from each of the measurements successively to obtain the best estimate of the track parameters. This procedure differs fundamentally from a standard χ^2 fit which is global in the sense that all measurements are considered at once. This necessitates that the procedure is started, which turns out to be slightly tricky and will not be discussed untill later. For the present discussion, it is assumed that the fit has been started and that a best estimate of the parameters, η, and a variance matrix, V, are obtained from the hits added so far. The procedure uses the next measurement to obtain a new best guess, η^{opt}, of the parameters and an updated variance matrix which contains information from the new measurement. When applied to track fitting, the track parameters must be transported to the location of the next measurement, before the hit can be added. The purpose of the transport is to account for the material. Trajectories going through material, where the particle may, for instance, multiple scatter, are simply accounted for by updating the variance matrix with the appropriate uncertainties. When a new measurement is added, the old estimate of the track parameters is de-weighted by the increase of the variance matrix due to the interactions with the material. This is discussed in detail in Section 9.5.2.

The Billoir fit is a χ^2 fit which accounts for the effects of multiple scattering and other sources of noise. This is accomplished without having to invert an $N \times N$ matrix. Instead, the Billoir fit adds the information from the measurements one at the time and extracts the current best guess together with a variance matrix. The estimate of the parameters, together with the variance matrix is then transported to the next measurement. If the next measurement is, for instance, a hit on a silicon wafer in a vertex detector, the parameters are transported to a point close to the measurement, and the variance matrix is updated with the effects of the material through which the track passed.

The fitting procedure consists conceptually of two parts, as depicted in Figure 9.1. The first part is to add the information from a measurement. This is simply described in the language of the weight matrix (also called the information matrix) which is the inverse of the variance matrix. It is convenient to think in terms of weight matrices since they are always defined. The variance matrix is singular if the system is under-constrained, i.e., if there are more free parameters than measurements.

The second part of the procedure is the concept of transport, in which the parameters and the variance matrix are transported between successive points. This must account for all physics processes that the particle is subjected to. This is most easily described in terms of the variance matrix, as will be evident below. This implies that the process of transporting and adding new measurements requires an inversion of a matrix in each step. This turns out to be fast as the inversions needed can be done analytically, using the so called Woodburry's matrix inversion theorem, described in the next section.

The steps of adding the information from a measurement and transporting the track parameters are described in detail in Sections 9.3 and 9.5 respectively.

9.3 Precise formulation of the Billoir fit

This section will provide a more technical presentation of the ideas described in the previous section.

For this discussion, assume that a set of track parameters, η, and their variance matrix, V_η, have been estimated, for example, from a earlier set of measurements, as shown in Figure 9.1. As described in the section above, the idea of the Billoir fit is to add one hit at the time. The next step is to understand how the estimated track parameters and their variance matrix

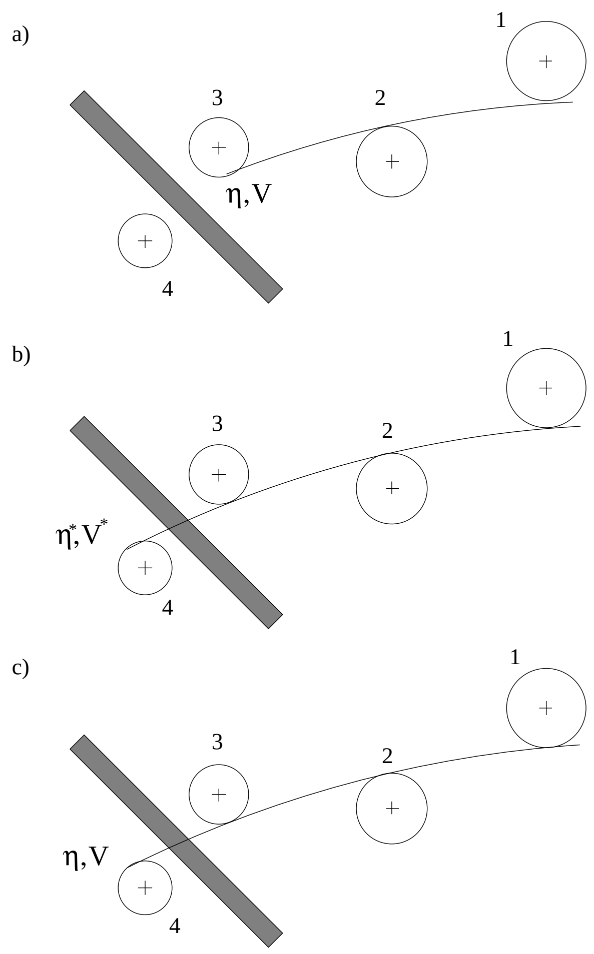

Figure 9.1: The idea of the Billoir fit. Part a) shows the track with the estimated parameters, η, and the variance matrix, V. This is after measurements 1, 2, and 3 have been added. b) shows the track after it has been transported to measurement 4. The track parameters, η^*, and the variance matrix, V^*, have been transported and updated with the effects of passing through material, shown as the shaded area. In c) the information from the last measurement, 4, has been added to the track and a new best estimate, η, is obtained together with an estimate of the variance matrix.

are affected by the addition of a new measurement.

After a hit is added a new set of track parameters, η^{opt} are obtained, with a variance matrix, V_η^{opt}. The hit provides a measurement of a quantity d^{meas}, e.g., a drift distance, with an estimated variance of σ^2. With, $d = d(\eta^{\text{opt}})$, the predicted value of the measured quantity as a function of the estimated track parameters. To proceed, form a combined χ^2 which combines the old estimate and the new measurement as

$$\chi^2 = \frac{\left(d(\eta^{\text{opt}}) - d^{\text{meas}}\right)^2}{\sigma^2} + \left(\eta^{\text{opt}} - \eta\right)^t V_\eta \left(\eta^{\text{opt}} - \eta\right). \tag{9.1}$$

Then new best estimate, η^{opt}, is obtained by minimizing the χ^2. Note that the new measurement, d^{meas}, is not correlated with η, which is estimated from the previously added measurements. To proceed it is convenient to write

$$\eta^{\text{opt}} = \eta + \tilde{\eta} \tag{9.2}$$

and to linearize $d(\eta^{\text{opt}})$ around η

$$d(\eta^{\text{opt}}) = d(\eta + \tilde{\eta}) = d(\eta) + \frac{\partial d}{\partial \eta}\tilde{\eta} = d(\eta) + D^t\tilde{\eta} \tag{9.3}$$

where

$$D = \frac{\partial d}{\partial \eta^t} \tag{9.4}$$

such that

$$\chi^2 = \frac{\left(d(\eta) - d^{\text{meas}} + D^t\tilde{\eta}\right)^2}{\sigma^2} + \tilde{\eta}^t V_\eta \tilde{\eta}. \tag{9.5}$$

In section 9.6 the calculation of the derivatives, D, are described for the different types of hits. Minimizing this χ^2 and solving for $\tilde{\eta}$ gives

$$\tilde{\eta} = \left(\frac{d^{\text{meas}} - d(\eta)}{\sigma^2}\right)\left(\frac{DD^t}{\sigma^2} + V_\eta^{-1}\right)^{-1} D. \tag{9.6}$$

It is now convenient to apply the so called Woodbury's matrix inversion formula

$$\left(\frac{DD^t}{\sigma^2} + V_\eta^{-1}\right)^{-1} = V_\eta - \frac{V_\eta D D^t V_\eta}{\sigma^2 + D^t V_\eta D}. \tag{9.7}$$

To show that this is correct is a simple exercise of multiplying together the two matrices. It is, however, a little more complicated to come up with the expression. This is a special case of a slightly more general formula [1]. Care has to be taken with numerical stability of this equation [2]. Using this, Eq. 9.7 can now be written as

$$\eta^{\mathrm{opt}} = \eta + \left(\frac{d^{\mathrm{meas}} - d(\eta)}{\sigma^2}\right)\left(V_\eta - \frac{V_\eta D D^t V_\eta}{\sigma^2 + D^t V_\eta D}\right) D. \tag{9.9}$$

As always, the variance matrix, $V_\eta^{\mathrm{opt}} = \langle \delta\eta^{\mathrm{opt}} \delta\eta^{\mathrm{opt}\,t} \rangle$, is needed. This is a straight forward, but tedious, calculation. Note that $\delta d(\eta) = D^t \delta\eta$. The result is, as expected

$$V_\eta^{\mathrm{opt}} = V_\eta - \frac{V_\eta D D^t V_\eta}{\sigma^2 + D^t V_\eta D}. \tag{9.10}$$

Writing this in terms of the weight matrix gives

$$W_\eta^{\mathrm{opt}} = W_\eta + \frac{D D^t}{\sigma^2}, \tag{9.11}$$

which shows that the information provided by the measurement is DD^t/σ^2. This is very natural and particularly easy to understand in the case when the measurement is a track parameter.

This successive addition of the measurements also allows for the construction of a χ^2. The total χ^2, is calculated by updating the χ^2 by the contribution from the addition of the last hit

$$\chi^2_{\mathrm{new}} = \chi^2_{\mathrm{old}} + \frac{(d(\eta) - d^{\mathrm{meas}} + D^t\tilde{\eta})^2}{\sigma^2} + \tilde{\eta}^t V_\eta \tilde{\eta}. \tag{9.12}$$

---

[1] $[A^{-1} + UV^t]^{-1} = A - A^{-t}U(I + V^t A U)^{-1} V^t A$

[2] To see when care has to be taken consider the case when the derivative D is zero for all parameters except the i:th. This happens when one of the track parameters is the measured quantity, e.g., when the track is referenced at the center of an axial wire the drift distance is the impact parameter. Then it is easy to see that

$$V^*_{\eta_{ii}} = V_{\eta_{ii}} - \frac{V_{\eta_{ii}} V_{\eta_{ii}}}{\sigma^2 + V_{\eta_{ii}}} = \left(\frac{1}{\sigma^2} + \frac{1}{V_{\eta_{ii}}}\right)^{-1} \tag{9.8}$$

The second equality shows that the new error on the i:th parameters is exactly that expected from averaging. However, if $\sigma^2 \ll V_{\eta_{ii}}$ inadequate numerical precision can cause the evaluation of $V_{\eta_{ii}}$ to give 0. This is a problem when a measurement with good precision (small σ^2) is to be added to a track that is not well measured (large $V_{\eta_{ii}}$), e.g., at the first addition of a z-measurement in the SVX. This is solved in the code by the use of double precision in this step.

The procedure described here gives the optimal track parameters at the last hit added to the track. To obtain the optimal estimate of the track parameters at other points along the track, the fit must be done in two passes, an inward pass and an outward pass and the results averaged at the points where the track parameters are needed. This procedure is called smoothing and is discussed in Section 9.8.

9.4 Starting the fit: linearization

So far the task of starting the fit has been avoided. Intuitively, the fit needs to be started with an initial variance matrix that has large errors on the initial track parameters. The errors have to be so large that the seed does not bias the final result. In principal the errors could be made very large, though not infinite since that would posses a numerical problem.

One catch with this approach, is that the fit might go astray, meaning that initial measurements pull the fit in a wrong direction. Instead of converging to the true parameters, almost anything might happen. There are two sources for this problem. The first is that the χ^2-estimator is an optimal estimator only for a linear problem, i.e., higher moments are neglected. Secondly, the transport will become more or less undefined when the track parameters to transport are not describing the real track. A typical scenario is when the first three measurements estimate a very large curvature, i.e., low momentum. Then the transport to and adding the next hit is not well defined, as shown in Figure 9.2. It has been suggested to start the fit with a variance matrix having a weight on the initial seed so that the first few measurements will not pull the fit astray. This has to be done with a great deal of care as the initial weight can not be so big that it affects the final fit, i.e., the seed should not bias the final result. This method is not used in this fitter. Instead, another approach mentioned by Fruhwirth [45] is used.

The approach taken is to linearize the fit around the seed track. This has two advantages. It linearizes the fit so that the χ^2 estimator becomes correct and it allows the transport to be made along the seed track, so that it will not be sensitive to the initial measurements if they pull the fit away from the true value.

In this linearization, the track is described as a perturbation, $\Delta\eta$, around a seed track, η_0,

$$\eta = \eta_0 + \Delta\eta. \tag{9.13}$$

It is important to correctly define the transport and how information is

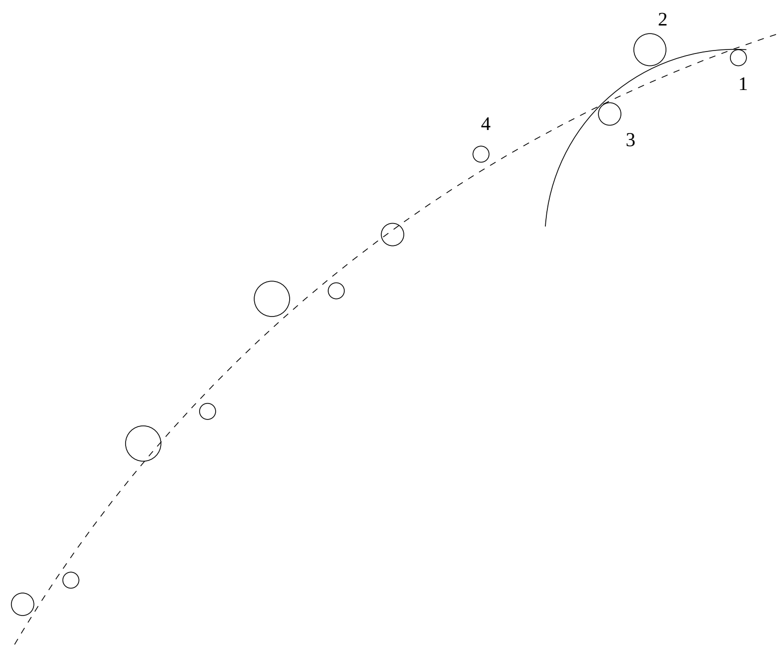

Figure 9.2: This figure illustrates a problem that can occur at the start of a fit. After the three first hits are added the estimate of the track parameters is not good. This might be caused by a hit that does not belong on the track or simply a hit with a large residual. After the first three hits are added the estimate of the track is indicated with the solid line. The actual track is indicated by the dashed line. When the track parameters are transported from the third hit to the fourth the transport is almost meaningless since the estimate of the track is far from what the actual track is. This is solved by linearizing the track around the seed.

added using the linearized track model.

First, the transport is discussed since it is straightforward. The seed track, η_0, is transported in the standard way to obtain η_0'. As usual the transport matrix, T, is defined as

$$T = \frac{\partial \eta_0'}{\partial \eta_0}. \tag{9.14}$$

Then $\Delta\eta$ is transported linearly as

$$\Delta\eta' = T\Delta\eta. \tag{9.15}$$

When a hit is added to the fit two things are needed: the predicted measurement, $d(\eta)$, from the track parameters as well as the derivatives of this measurement with respect to the track parameters. The predicted measurement, $d_0 = d(\eta_0)$, from the seed track is calculated and the derivatives, D, of this distance with respect to the seed track parameters are calculated. The predicted measurement of the track is then evaluated as

$$d = d_0 + D^t\Delta\eta. \tag{9.16}$$

When the track is traced through the material, the path and the current momentum of the particle are taken from the seed track. This very simple procedure has shown to be stable against the problems discussed above. The initial error matrix can be arbitrarily large.

9.5 Track parameter transport

This section discusses how track parameters and their variance matrix are transported. The transport is divided into two steps, the first being simply a change of variables. If a track is described by a set of track parameters, η, at a particular point, it will be described by another set of parameters, $\eta' = f(\eta)$, at another point on the track. It is then easy to express the variance matrix at the new point by a simple change of variables. Exactly what is meant by describing the track at a point is discussed below.

The second part of the transport is to update the track parameters and the error matrix with the effects from interactions with the material the particle passes through. The interactions can be divided in two categories. The first is multiple (coulomb) scattering, MS. This does not affect the estimated track parameters after transport, it only contributes to the error matrix. The second category is energy loss, for example due to ionization, which changes the track parameters. This also affects the error matrix, not only from the effect of straggling in the energy loss, but also by a more subtle effect which occurs when the curvature is changed.

So far a particular representation of the track has not been needed. However, below when describeing the effects of the multiple scattering and energy loss it is necessary to use a particular representation. In Appendix A the track parametrization used in this implementation of the Billoir fitter is described.

9.5.1 Transport in vacuum

This part of the transport is in some sense trivial, but is still worth describing in detail. The track, η, with variance matrix, V, is given at a point, x. The point x should be thought of as the origin of a local coordinate system in which the track is described, see Figure 9.3. The local coordinate system is simply a translation of the global coordinate system. The operation of transporting a set of track parameters is defined as the operation of finding the new track parameters, η', with variance matrix, V', at a new reference point, x'. This formulation of the problem of track parameter transport is slightly different from what is used elsewhere, where, for example, a transport is made a certain distance along the track or to the interception of the track with some given surface. The approach taken here is conceptually very simple and is very useful is the implementation of the Billoir fitter.

The transport routines described here work in a uniform magnetic field along the z-axis, such that the track forms a helix in space. However, this approach can be modified to handle nonuniform magnetic fields, see Section 9.7.

The track parameters at the new reference point, x', can be written as

$$\eta' = f(\eta). \tag{9.17}$$

Defining the transport matrix, T, by

$$T_{ij} = \frac{\partial \eta'_i}{\partial \eta_j}, \tag{9.18}$$

from this it is easy to find V'.

$$V'_{ij} = \langle \delta\eta'_i \delta\eta'_j \rangle = \langle T_{ik}\delta\eta_k T_{jl}\delta\eta_l \rangle = T_{ik}V_{kl}T_{jl} \tag{9.19}$$

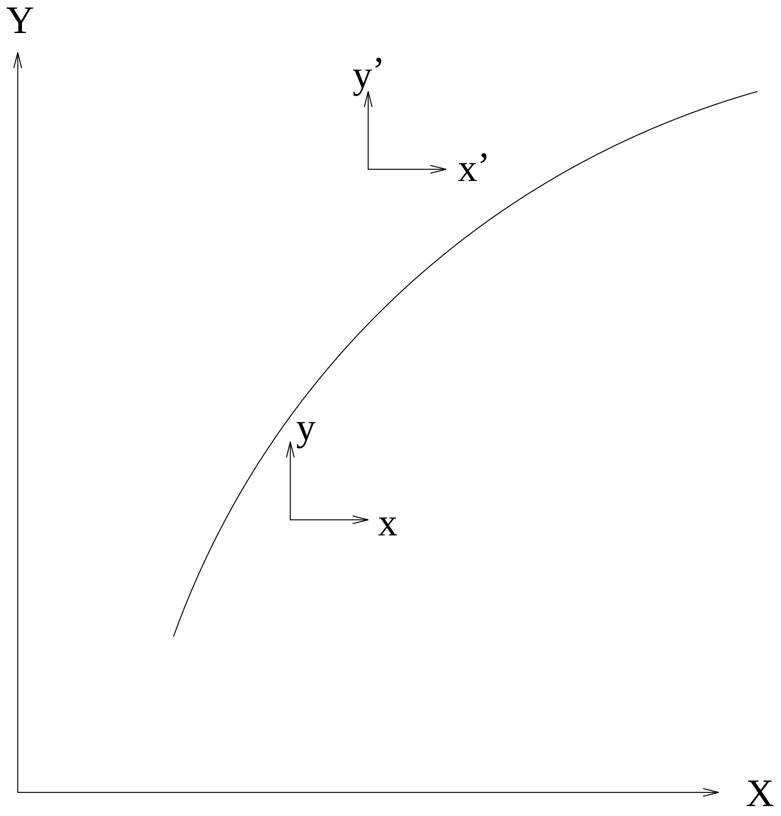

Figure 9.3: The different coordinate systems used in the transport. The XY coordinate system is the global coordinate system. The track parameters, η, are originally given in the xy coordinate system. In the transport operation, the track is described in the x'y' coordinate system. That is the track parameters are always given at the point of closest approach to the (local) origin in the xy-plane. When material is present, i.e., transport is no longer in vacuum, the transported track parameters will also be modified by the effects of the material, and not only the geometrical effect of describing the same circle at different points.

or, in matrix form, simply

$$V' = TVT^t. \tag{9.20}$$

The transport is exact in the sense that there are no approximations made about the transport distance being small compared to the radius of curvature. Note also that there are only 10 nontrivial elements in T, see Appendix B.

9.5.2 Multiple scattering

The treatment of multiple scattering makes the assumption that the scattering occurs in a thin plane at the point of closet approach of the track to the new reference point. The scattering is assumed to be gaussian with a width of θ_0 in the scattering plane. θ_0 is calculated using formulas given in [36]. It is obvious that the scattering will affect the variance matrix elements $V_{\phi\phi}$ and V_{tt}, which describe the direction of the track, as MS adds an uncertainty in the direction. However, it does not change the momentum of the track. As the total momentum of the track depends on t and K, there will also be contributions from MS to the variance matrix elements V_{KK} and $V_{\mathrm{K}t}$. The rest of this section shows this in detail.

Figure 9.4 shows the geometry needed to understand how multiple scattering effects the track parameters. The vector, v, is a unit vector pointing along the direction of the track. The two angles, ϕ and λ, describe the direction of the track. ϕ is simply one of the track parameters, while λ is related to t by $t = \tan\lambda$. Multiple scattering is described by two uncorrelated scatterings with r.m.s. of θ_0 in two perpendicular planes containing the vector v, see [46]. As the first plane, we chose the plane spanned by the vectors v and $\hat{z}$. It is obvious that a scattering in this plane only effects $V_{\lambda\lambda}$ [3]

$$V_{\lambda\lambda} \rightarrow V_{\lambda\lambda} + \theta_0^2 \tag{9.21}$$

or expressed in terms of the track parameter t,

$$V_{tt} \rightarrow V_{tt} + (1+t^2)^2\theta_0^2. \tag{9.22}$$

The other scatter plane can be described as the plane perpendicular to the vector r, see Figure 9.4. A rotation an angle θ_0 around r corresponds to

[3] Given an angle ψ with a variance matrix $V_{\psi\psi}$, $V_{\psi\psi} = \langle(\psi - \bar{\psi})^2\rangle$. If this angle is subject to an change of direction by an angle θ, such that the mean is zero and the rms is θ_0, like in the case of MS, then with $\phi' = \phi + \theta$ it is easy to see that $V_{\phi'\phi'} = V_{\phi\phi} + \theta_0^2$ when θ, the MS, is uncorrelated with ϕ.

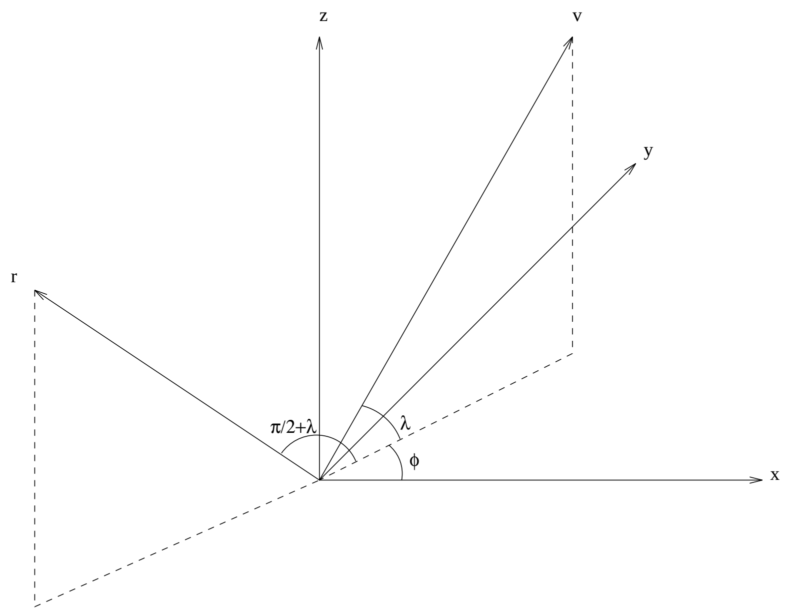

Figure 9.4: Geometry of multiple scattering. The direction of the track is indicated by the vector v in the figure. The multiple scattering is described as two independent scatterings in two perpendicular planes that both contains the vector v. As the first plane, take the plane that is spanned by v and the z axis. The second plane is taken as the plane that is perpendicular to the vector r in the figure. r lies in the first plane and is perpendicular to v.

a scattering θ_0 in the second plane. The goal is to find how this affects ϕ (λ is, as will be shown below, unchanged under the infinitesimal rotation considered). One way of obtaining the desired result is to note that under a rotation, θ_0, around the r axis $v \to v + dv$ where

$$dv = \theta_0 v \times r. \tag{9.23}$$

Using

$$\begin{aligned} v &= (\cos\phi\cos\lambda, \sin\phi\cos\lambda, \sin\lambda), & (9.24)\\ r &= (-\cos\phi\sin\lambda, -\sin\phi, \sin\lambda, \cos\lambda) & (9.25) \end{aligned}$$

gives

$$dv = \theta_0(-\sin\phi, \cos\phi, 0). \tag{9.26}$$

From this it is deduced that

$$\cos\phi \to \cos\phi - \frac{\theta_0}{\cos\lambda}\sin\phi = \cos(\phi + \frac{\theta_0}{\cos\lambda}) \tag{9.27}$$

which implies that

$$V_{\phi\phi} \to V_{\phi\phi} + \frac{\theta_0^2}{\cos^2\lambda} = V_{\phi\phi} + \theta_0^2(1+t^2). \tag{9.28}$$

This is the first part of the treatment of multiple scattering. There is another effect to be considered, stemming from the fact that the momentum, or more importantly the error in the momentum, is not affected by multiple scattering. To see how this comes about consider the momentum as a function of the track parameters

$$P = \frac{A}{|\mathrm{K}|}\sqrt{1+t^2}. \tag{9.29}$$

Differentiating this gives

$$dP = -d\mathrm{K}\frac{P}{\mathrm{K}} + dt\frac{Pt}{1+t^2}. \tag{9.30}$$

Momentum is conserved under MS, meaning $dP = 0$, which gives

$$d\mathrm{K} = dt\frac{\mathrm{K}t}{1+t^2}. \tag{9.31}$$

From this it is easy to see that the following elements in the variance matrix also need to be updated

$$\begin{aligned} V_{\mathrm{K}t} &\to V_{\mathrm{K}t} + \mathrm{K}t(1+t^2)\theta_0^2, & (9.32)\\ V_{\mathrm{KK}} &\to V_{\mathrm{KK}} + \mathrm{K}^2t^2\theta_0^2. & (9.33) \end{aligned}$$

To summarize there are 4 elements, $V_{\phi\phi}$, V_{tt}, $V_{\mathrm{K}t}$, and V_{KK} that must be updated due to the uncertainty caused by multiple scattering.

9.5.3 Energy loss

When the particle passes through material it loses some amount of energy, ΔE, which affects its trajectory. There is straggling in the energy loss–it is not deterministic. Let δE denote the r.m.s. of the straggling in ΔE, $\delta E = \sqrt{\langle(\Delta E - \overline{\Delta E})^2\rangle}$. The first thing to do is to update the curvature, K, with the effect of the energy loss, which can be written in the following way

$$\mathrm{K}' = \mathrm{K}\frac{P}{\sqrt{P^2 + 2E\Delta E + \Delta E^2}}. \tag{9.34}$$

The other effect that needs to be incorporate is the straggling. This adds a contribution to the variance matrix element V_{KK}. Using

$$|\mathrm{K}| = \frac{A\sqrt{1+t^2}}{\sqrt{E^2 - m^2}} \tag{9.35}$$

it is easy to find that

$$V_{\mathrm{KK}} \to V_{\mathrm{KK}} + \frac{\mathrm{K}^2 E^2}{P^4}\delta E^2. \tag{9.36}$$

However, there is yet one more effect from the change of curvature that is important for low momentum tracks. When a track changes its curvature the variance matrix is also affected. This is a geometrical effect and is not related to the effects of straggling. To work this out consider the energy loss as a change of variables

$$\eta \xrightarrow{\Delta E} \eta' \tag{9.37}$$

where the primed and unprimed variables are related by

$$\begin{aligned}
\mathrm{K}' &= \mathrm{K}\frac{P}{\sqrt{P^2 + 2E\Delta E + \Delta E^2}}, && (9.38)\\
\phi' &= \phi, && (9.39)\\
d' &= d, && (9.40)\\
t' &= t, && (9.41)\\
z' &= z. && (9.42)
\end{aligned}$$

There are only two nontrivial derivatives,

$$\begin{aligned}
a &\equiv \frac{\partial \mathrm{K}'}{\partial \mathrm{K}} = \frac{\mathrm{K}'^3}{\mathrm{K}^3}\frac{E'}{E} && (9.43)\\
b &\equiv \frac{\partial \mathrm{K}'}{\partial t} = \frac{t}{1+t^2}\mathrm{K}'(1 - \frac{\mathrm{K}'^2}{\mathrm{K}^2}\frac{E'}{E}). && (9.44)
\end{aligned}$$

This gives

$$T = \frac{\partial \eta'}{\partial \eta} = \begin{bmatrix} a & 0 & 0 & b & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9.45}$$

The new variance matrix, V', given by $V' = TVT^t$, differs from V in the following elements

$$V'_{\mathrm{KK}} = a^2 V_{\mathrm{KK}} + 2ab V_{\mathrm{K}t} + b^2 V_{tt}, \tag{9.46}$$
$$V'_{\mathrm{K}\phi} = a V_{\mathrm{K}\phi} + b V_{t\phi}, \tag{9.47}$$
$$V'_{\mathrm{K}d} = a V_{\mathrm{K}d} + b V_{td}, \tag{9.48}$$
$$V'_{\mathrm{K}t} = a V_{\mathrm{K}t} + b V_{tt}, \tag{9.49}$$
$$V'_{\mathrm{K}z} = a V_{\mathrm{K}z} + b V_{tz}. \tag{9.50}$$

For a low momentum track, where 10% of the momentum is lost $\mathrm{K}'/\mathrm{K} = 0.9$ and $V'_{\mathrm{KK}} \approx 0.9^6 V_{\mathrm{KK}} = 0.53 V_{\mathrm{KK}}$. This makes this effect important for low momentum, while it is completely negligible for high momentum.

9.6 Adding information to the fit

In Section 9.3 the addition of information, hits, to the track fit was described. It was assumed that the hit provides a measurement, d, of some quantity, e.g., a drift distance or a position on a silicon wafer. The addition of this information was done in an abstract form in which the derivative, D, see Eq. 9.4, was needed. The purpose of this section is to give the explicit form of the derivatives, D, for the different types of hits that are considered.

9.6.1 Axial wires

When treating axial wires, the derivatives needed are trivial, as it is assumed that the track parameters are expressed in a coordinate system where the wire is at the origin. Therefore the projected drift distance from the track, $d^{\min}(\eta)$, is d (one of the track parameters). This leads to only one non-zero derivative, which is trivial,

$$\frac{dd^{\min}}{dd} = 1. \tag{9.51}$$

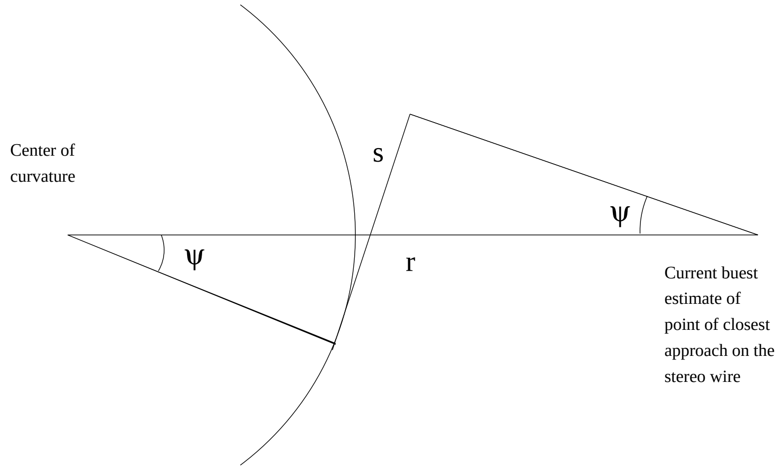

Figure 9.5: This figure shows how the track, the arc, is approximated with a straight line and how the distance s is found. Note that when the point is close to the track compared to the radius the distance r is well approximated by the radius of curvature. However, when the point is far away, for example outside the arc as in the figure this is not a good approximation. Under normal circumstances the point, a given wire, is close to the track since the drift distance is much smaller than the curvature.

9.6.2 Stereo wires

The stereo wire hits are more involved than the axial wire. The first problem is that the point of closest approach between the track and the wire has to be found. The solution to this problem can not be obtained in closed form as it involves a transcendental equation. Instead, the solution is found numerically by an iterative process, which converges in few iterations with a reasonable initial guess. The stereo wires are parametrized by s in the following way

$$x_w = x_{0w} + c_x s, \tag{9.52}$$
$$y_w = y_{0w} + c_y s, \tag{9.53}$$
$$z_w = s \tag{9.54}$$

such that the vector $(c_x, c_y, 1)$ is along the direction of the stereo wire and x_{0w} and y_{0w} are the coordinates of the stereo wire in the $z = 0$ plane. The track is parametrized by the turning angle, ψ,

$$x_t = -d \sin\phi + \frac{1}{2\mathrm{K}} \left(\sin(\phi + q\psi) - \sin\phi\right), \tag{9.55}$$

$$y_t = d\cos\phi + \frac{1}{2\mathrm{K}}\left(\cos\phi - \cos(\phi + q\psi)\right), \tag{9.56}$$

$$z_t = z + \frac{tq\psi}{2\mathrm{K}}. \tag{9.57}$$

The goal is to find the ψ that minimizes the distance between the track and the wire. As the first guess of the point of closest approach $\psi = 0$ is taken. To iterate this, the track is approximated as a straight line at the point (x_t, y_t, z_t). Using vector notation, the wire and the track are parametrized as

$$x_w = x_{0w} + v_w s_w, \tag{9.58}$$

$$x_t = x_{0t} + v_t s_t, \tag{9.59}$$

where $x_{0t} = (x_t, y_t, z_t)$ is the current best guess of the closest point on the track and

$$v_w = (c_x, c_y, 1)/\sqrt{c_x^2 + c_y^2 + 1}, \tag{9.60}$$

$$v_t = (\cos(\phi + q\psi), \sin(\phi + q\psi), t)/\sqrt{1 + t^2}. \tag{9.61}$$

The next step is to find s_t and s_w which minimize the distance

$$\Delta = \sqrt{(x_w - x_t)^2} = \sqrt{(x_{0w} + v_w s_w - x_{0t} - v_t s_t)^2}. \tag{9.62}$$

Solving the above equation for the s_t which minimizes Δ gives

$$s_t = \frac{(x_{0w} - x_{0t}) \cdot (v_w(v_w \cdot v_t) - v_t)}{(v_w \cdot v_t)^2 - 1}, \tag{9.63}$$

where use has been made of the fact that v_w and v_t are unit vectors. This gives a distance, s_t, to move along the track to get a new estimate of the point on the track closest to the wire, see Figure 9.5.

$$\psi \to \psi + \frac{s_t}{r\sqrt{1 + t^2}} \tag{9.64}$$

With this new estimate of ψ, the steps above are repeated untill the procedure has converged. The convergence criteria is that the estimated distance to the point of closest approach is less than 10 μm.

After having found the point of closest approach, expressed as ψ, the derivatives of the distance of closest approach to the stereo wire with respect to the track parameters can now easily be evaluated, without making any approximations. It is convenient to work out the derivatives with respect to $d^2_{\min}$ and use

$$\frac{dd_{\min}}{d\eta} = \frac{1}{2d_{\min}}\frac{dd^2_{\min}}{d\eta}. \tag{9.65}$$

The next step will be very explicit, to show how the derivatives work out. The end result is just what's expected. First write $d^2_{\min}$ with all dependencies on η

$$\begin{aligned}
d^2_{\min} &= \left[x_t\left(\eta, \psi_w^{\min}(\eta)\right) - x_{0w} - c_x s_{\min}(\eta)\right]^2 \\
&+ \left[y_t\left(\eta, \psi_w^{\min}(\eta)\right) - y_{0w} - c_y s_{\min}(\eta)\right]^2 \\
&+ \left[z_t\left(\eta, \psi_w^{\min}(\eta)\right) - s_{\min}(\eta)\right]^2 \\
&= \Delta x^2 + \Delta y^2 + \Delta z^2.
\end{aligned} \tag{9.66}$$

Differentiating this with respect to η gives

$$\begin{aligned}
\frac{dd^2_{\min}}{d\eta} &= 2\Delta x\left(\frac{\partial x_t}{\partial \eta} + \frac{\partial x_t}{\partial \psi_{\min}}\frac{d\psi_{\min}}{d\eta} - c_x\frac{ds_{\min}}{d\eta}\right) \\
&+ 2\Delta y\left(\frac{\partial y_t}{\partial \eta} + \frac{\partial y_t}{\partial \psi_{\min}}\frac{d\psi_{\min}}{d\eta} - c_y\frac{ds_{\min}}{d\eta}\right) \\
&+ 2\Delta z\left(\frac{\partial z_t}{\partial \eta} + \frac{\partial z_t}{\partial \psi_{\min}}\frac{d\psi_{\min}}{d\eta} - \frac{ds_{\min}}{d\eta}\right) \\
&= 2\Delta x\frac{\partial x_t}{\partial \eta} + 2\Delta x\frac{\partial x_t}{\partial \eta} + 2\Delta x\frac{\partial x_t}{\partial \eta}
\end{aligned} \tag{9.67}$$

since the conditions for the minimum are

$$0 = \Delta x\frac{\partial x_t}{\partial \psi_{\min}} + \Delta y\frac{\partial y_t}{\partial \psi_{\min}} + \Delta z\frac{\partial z_t}{\partial \psi_{\min}}, \tag{9.68}$$

$$0 = \Delta x c_x + \Delta y c_y + \Delta z. \tag{9.69}$$

The needed derivatives, e.g., $\partial x_t/\partial\eta$ are easily computed from Eq. 9.55-9.57.

9.6.3 Cathodes

As for the axial wires this is a trivial derivative. The measurement is a z-position so $dd/dz = 1$ and all other derivatives are zero. This assumes that the track is transported to the radius of the measurement.

9.6.4 Charge division hits

Treated in the same way as the cathodes.

9.6.5 Silicon strip hits

The calculation of the derivatives for a silicon strip measurement is similar to that for a stereo wire. It proceeds in two steps. The first step is to find the point of intercept of the track with the wafer. This is done with an iterative process very similar to finding the distance of closest approach to a stereo wire. Once this point is found, it is easy to write down, exact expressions for the derivative with respect to the track parameters. No assumptions are made about the wafers being parallel to the z-axis.

The iteration for finding the intercept, described in terms of ψ, is done by approximating the track as a straight line. With the notation in Figure 9.6, the condition for the intercept with the plane of the wafer is

$$n \cdot (x_0 - (x_t + sv)) = 0. \tag{9.70}$$

The plane of the wafer is described by its normal, n, and a point in the plane, x_0. x_t is the current guess of the position on the track for the intercept of the track and the wafer and v is the tangent of the track at that point. It is easy to solve this equation for the distance, s, along the track to the intercept,

$$s = \frac{n \cdot (x_o - x_t)}{n \cdot v}. \tag{9.71}$$

This leads to the following change in ψ

$$\psi \to \psi + \frac{2s\mathrm{K}}{\sqrt{1+t^2}}. \tag{9.72}$$

This procedure is repeated untill the intercept is found with sufficient precision, 10 μm.

After the intercept of the track and the wafer is found in terms of the turning angle, ψ_0, the derivatives needed are easily evaluated. With u being a unit vector in the silicon wafer plane along the direction where the measurement, d, is made, the predicted measurement d_t is easily found as

$$d_t = u \cdot (x_0 - x_t). \tag{9.73}$$

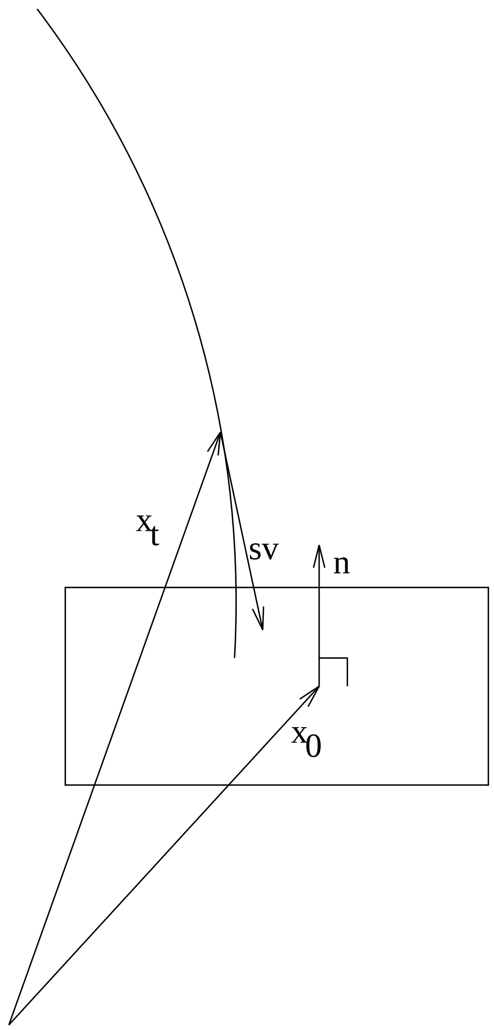

Figure 9.6: This figure shows how the track, the arc, is approximated with a straight line and how the distance s is to the plane is found.

The derivatives needed are

$$\frac{dd_t}{d\eta} = u \cdot \frac{dx_t}{d\eta} = u \cdot \left(\frac{\partial x_t}{\partial \eta} + \frac{d\psi_0}{d\eta} \frac{\partial x_t}{\partial \psi_0} \right). \tag{9.74}$$

The only derivative that is somewhat tricky to evaluate is $d\psi_o/d\eta$. The easiest way to evaluate it is to use

$$0 = n \cdot \frac{dx_t}{d\eta}(\eta, \psi_0(\eta)) = n \cdot \frac{\partial x_t}{\partial \eta} + \frac{d\psi_0}{d\eta} n \cdot \frac{\partial x_t}{\partial \eta}. \tag{9.75}$$

This gives

$$\frac{d\psi_0}{d\eta} = -\frac{n \cdot \dfrac{\partial x_t}{\partial \eta}}{n \cdot \dfrac{\partial x_t}{\partial \psi_0}}. \tag{9.76}$$

All other derivatives are trivial to evaluate.

9.7 Nonuniform magnetic field

So far the discussion has been concerned with fits of tracks in a uniform magnetic field along the z-axis. This is what the code was designed to do. However, there is a mechanism to allow for nonuniform magnetic fields. The idea is to treat the difference in the magnetic field,

$$\Delta B = B_{\text{true}} - B_{\text{uniform}}, \tag{9.77}$$

as a perturbation. The fit is performed as if the field was uniform along the z-axis, and the residual magnetic field is used to correct the transport. This is explained carefully below. One effect this has is that when extracting the momentum of the particle, say at the origin, the uniform magnetic field is used, not the actual magnetic field, i.e., the track parameter K is the curvature of the track with the current momentum in a magnetic field with the uniform strength not the local field. This might seem somewhat counter intuitive at first but is straight forward to implement.

Starting from the well known formula for the force on a charged particle, with charge q (in unit charges), in a magnetic field B (Tesla)

$$F = 0.3qB \times v \tag{9.78}$$

where v is the particles velocity in m/s and the force is given in units of GeV/s. A very useful expression is obtained after integrating the above equation with respect to time

$$P = 0.3qB \times x. \tag{9.79}$$

This equation describes how the momentum of a particle is changing as it goes a distance x in a magnetic field B. The equation is meaningful as long as the distance, x, in the magnetic field can be approximated with a straight line. Note that a global form of this is the familiar expression $P = 0.3qB\rho$ which gives the (transverse) momentum of a particle with a radius of curvature ρ.

The interest here is to calculate the residual effect of the magnetic field ΔB as a particle goes a distance x.

The residual magnetic field gives the following change in the momentum of the particle

$$\Delta P = 0.3q\Delta B \times x. \tag{9.80}$$

The next step is to correct the track parameters for this change. The track parameters affected by this are K, ϕ, and t. The changes ΔK, $\Delta\phi$, and Δt to correct for the nonuniformity in the magnetic field will now be calculated. Writing down the changes of a particle's momentum with respect to the changes in the track parameters and equate them with the calculated changes from the residual magnetic field give

$$\Delta P_x = -\frac{\Delta K}{|K|}P_x - \Delta\phi P_y, \tag{9.81}$$

$$\Delta P_y = -\frac{\Delta K}{|K|}P_y + \Delta\phi P_x, \tag{9.82}$$

$$\Delta P_z = \frac{\Delta K}{|K|}P_z + \Delta t\frac{A}{|K|}. \tag{9.83}$$

Solving these equations gives

$$\Delta\phi = \frac{|K|}{A}(\Delta P_y \cos\phi - \Delta P_x \sin\phi), \tag{9.84}$$

$$\Delta K = \frac{|K|K}{A}(\Delta P_x \cos\phi + \Delta P_y \sin\phi), \tag{9.85}$$

$$\Delta t = \Delta P_z\frac{|K|}{A} + \frac{t\Delta K}{K}. \tag{9.86}$$

This is easiest to derive using the constraint $\Delta P \cdot P = 0$. The track is then updated with these corrections. The magnetic field, ΔB, is evaluated at the midpoint of the transport step.

9.8 Smoothing

After a track fit, the optimal track parameters at the point of closest approach to the origin are obtained. However, only a poor estimate of the track parameters at any other point along the track is obtained. As the fit is performed towards the innermost point the best estimate is successively obtained for the track parameters *based on the hits so far encountered.* To obtain the optimal track parameters at any other point along the track based on all hits on the track the fit has to be run from both ends of the track to a given point and the results of the two fits averaged. This procedure is know as smoothing [45]. A practical way of doing this is to save the track parameters from the inward pass of the fit and then on the second, outward, pass average the estimate, which is based on the hits inside the point, with the first estimate at that point from the inward pass. It is important to realize that these two estimates are uncorrelated even in the presence of multiple scattering.

Given two estimates, η_1 and η_2, for the same quantity η with the corresponding variance matrices V_1 and V_2. η_1 and η_2 are assumed to be uncorrelated. To find the best estimate of η write down the likelihood as a function of η

$$\mathcal{L} = e^{-\frac{1}{2}(\eta-\eta_1)^t V_1^{-1}(\eta-\eta_1) - \frac{1}{2}(\eta-\eta_2)^t V_2^{-1}(\eta-\eta_2)}. \tag{9.87}$$

and take a derivative with respect to η^t and set it to zero

$$0 = \frac{\partial(-ln\mathcal{L})}{\partial\eta^t} = V_1^{-1}(\eta-\eta_1) + V_2^{-1}(\eta-\eta_2). \tag{9.88}$$

Solving this for η gives

$$\eta = (V_1^{-1} + V_2^{-1})^{-1}(V_1^{-1}\eta_1 + V_2^{-1}\eta_2). \tag{9.89}$$

As always, the variance matrix for η is needed, it is found simply from the definition

$$V_\eta = \langle\delta\eta\delta\eta^t\rangle = ... = (V_1^{-1} + V_2^{-1})^{-1}. \tag{9.90}$$

Eq. 9.89 can also be rewritten in the following way

$$\eta = V_2(V_1+V_2)^{-1}\eta_1 + V_1(V_1+V_2)^{-1}\eta_2. \tag{9.91}$$

This shows that only $V_1 + V_2$ needs to be invertible to allow for the construction of the average.

9.9 Hit residuals

This section is devoted to a discussion of the residuals calculated in the fit and their distribution, i.e., their errors. This is important since if the residual distributions are not understood, the derived quantities, like the χ^2-distribution and the variance matrix of the track parameters, are also not understood. The hit residuals are also used internally in the fitter if hit deletion is used. Hits with too large residuals are deleted.

The fitter can calculate two types of hit residuals; where the hit is included in the track fit, and without the hit included. First, the case when the hit is *not* included in the fit is discussed, as this is the simple case. The residual with the hit included is somewhat more complicated, though the final result is very simple.

When the hit is not included in the fit, the projected measurement, d_{track}, from the track is uncorrelated with the measurement, d_{meas}, and the residual,

$$r \equiv d_{\text{meas}} - d_{\text{track}}, \tag{9.92}$$

have the variance

$$\sigma_r^2 = \sigma_{d_{\text{meas}}}^2 + \sigma_{d_{\text{track}}}^2. \tag{9.93}$$

When the hit is included in the track fit the residual and the projected measurement from the track becomes correlated and Eq. 9.93 is no longer correct. The easiest way to find the variance is to first combine the measurement, d_{meas}, with the projected track without the hit included , d_{track}. Averaging these uncorrelated estimates give

$$d_{\text{c}} = (\frac{d_{\text{track}}}{\sigma_{d_{\text{track}}}^2} + \frac{d_{\text{meas}}}{\sigma_{d_{\text{meas}}}^2})\sigma_{\text{c}}^2, \tag{9.94}$$

where

$$\frac{1}{\sigma_{\text{c}}^2} = \frac{1}{\sigma_{d_{\text{track}}}^2} + \frac{1}{\sigma_{d_{\text{meas}}}^2}. \tag{9.95}$$

σ_{c} is the variance of d_{c}, the estimated projected drift distance for a hit where the hit is included in the track fit. Defining the residual r by

$$r \equiv d_{\text{meas}} - d_{\text{c}} = d_{\text{meas}} - (\frac{d_{\text{track}}}{\sigma_{d_{\text{track}}}^2} + \frac{d_{\text{meas}}}{\sigma_{d_{\text{meas}}}^2})\sigma_{\text{c}}^2, \tag{9.96}$$

it is now easy to find the variance on the residual

$$\sigma_r^2 = \langle \delta r \delta r \rangle = ... = \sigma_{d_{\text{meas}}}^2 - \sigma_{\text{c}}^2. \tag{9.97}$$

This shows that the variance of the hit residuals is simply given by the variance of the resolution of the measurement minus the variance of the projected track including the hit. Note that the difference between this case and the case were the hit is left out of the track fit is just the sign. σ_r^2 is always positive since σ_c^2 contains the information from the hit, and therefore is smaller than $\sigma_{d_\mathrm{meas}}^2$.

Chapter 10

Implementation and performance

10.1 Implementation

The implementation of the Billoir fitter is describe in detail in [47]. Here, only a brief discussion of the implementation is presented. One of the concerns in the implementation was speed. Quite early in the project it was realized that inefficient code could lead to very poor performance. However, in optimization there is always a trade-off between speed and simplicity in coding. It is also highly desirable to keep the code as modular as possible to make changes in the organization of the code easy. Most of the modifications to make the code efficient have been made in specialized routines, e.g., matrix inversion and the multiplication of the matrices TVT^t for error matrix transport. A few other things that has been made to make the code efficient will be discussed below.

The code, written in Fortran 77, contains approximately 70 subroutines. The fitter is logically divided into two parts. The first part is the core of the fitter that performs the operations described in the previous sections, e.g., track parameter transport and addition of hits to the fit. The second part is the material interface, used in the transport. This part contains routines for tracing the particle through the detector elements. This code can trace a straight line through the detector and returns the distances the particle went in the various materials. This is used to calculate the multiple scattering and the energy loss of the particles.

The standard fit, performed on the real data, consists of fitting a set of hits, obtained from the pattern recognition, to obtain the optimal track parameters at the origin for the five particle hypothesis, e, μ, π, K, and p. Since one of the most time consuming operations is to trace the particles through the material in the detector the material the track went through in the first fit is saved and reused for the subsequent fits.

The CLEO II pattern recognition is quite generous in which hits are accepted on tracks. Therefor, the fitter also has to perform hit deletion. This is only done on the first particle hypothesis being fitted, the pion. The fit is performed both inward and outward for the pion hypothesis and the residuals are calculated at every hit. Hits which are more than three standard deviations from the projected track are deleted unless there are to few hits on the track. The maximum number of hits that are deleted is three. The hit deletion is a quite time consuming operation since it may require as much as three complete refits of the track.

10.2 Performance

The code has been tested extensively both on Monte Carlo and data [48]. The advantage of testing the code on Monte Carlo is that it is a controlled environment in which the code can be tested accurately. The tests on Monte Carlo have shown that the code correctly can handle the effects of energy loss and multiple scattering. In Figure 10.1 the track parameters for 1000 events are shown. The plots show the reconstructed minus the generated–true–track parameters divided by the error from the fit. This distribution should have unit r.m.s. and be gausian since the resolution function and the multiple scattering in the Monte Carlo is gausian. The sixth plot shows the probability of χ^2 distribution which should be flat.

At the time of writing this dissertation the Billoir fitter is just about to be used in the re-compress of the CLEO II data set. A small fraction of the data set has already been processed and in Figure 10.2 the $D^* - D$ mass difference is shown using the Billoir fitter. The lower plots show the same quantity using the old fitter, TF3FIT. The plots to the right have a vertex constraint imposed, i.e., that the kaon and the two pions come from the same production point. It is clear that the vertex constraint help the data fitted with the Billoir fitter while the data fitted with TF3FIT is not helped by this constraint.

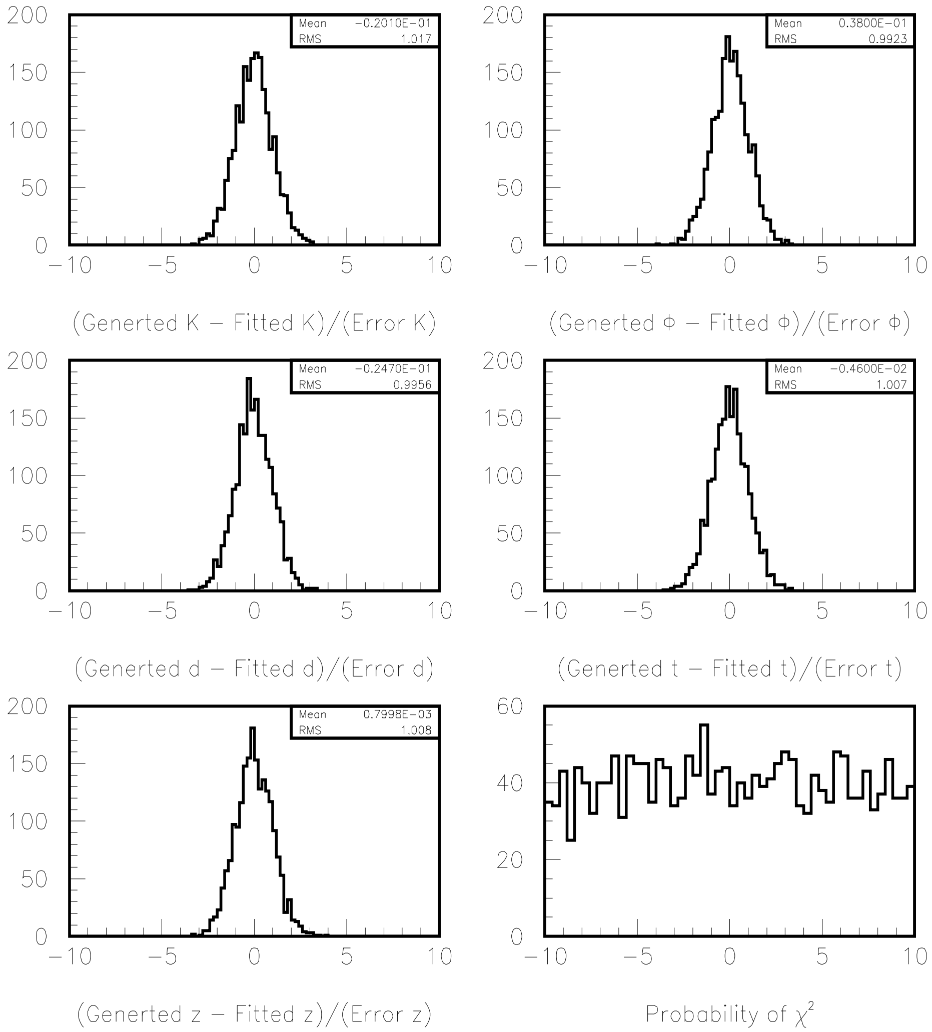

Figure 10.1: Track parameter residuals in Monte Carlo. The Monte Carlo is a truth table Monte Carlo, i.e, all hits on the track are correct. The mean of the residual distributions should be zero and the width, the r.m.s., should be one. The sixth plot shows the probability of χ^2 distribution which should be flat. These plots show that the fitter is correctly handling the effects of multiple scattering and energy loss.

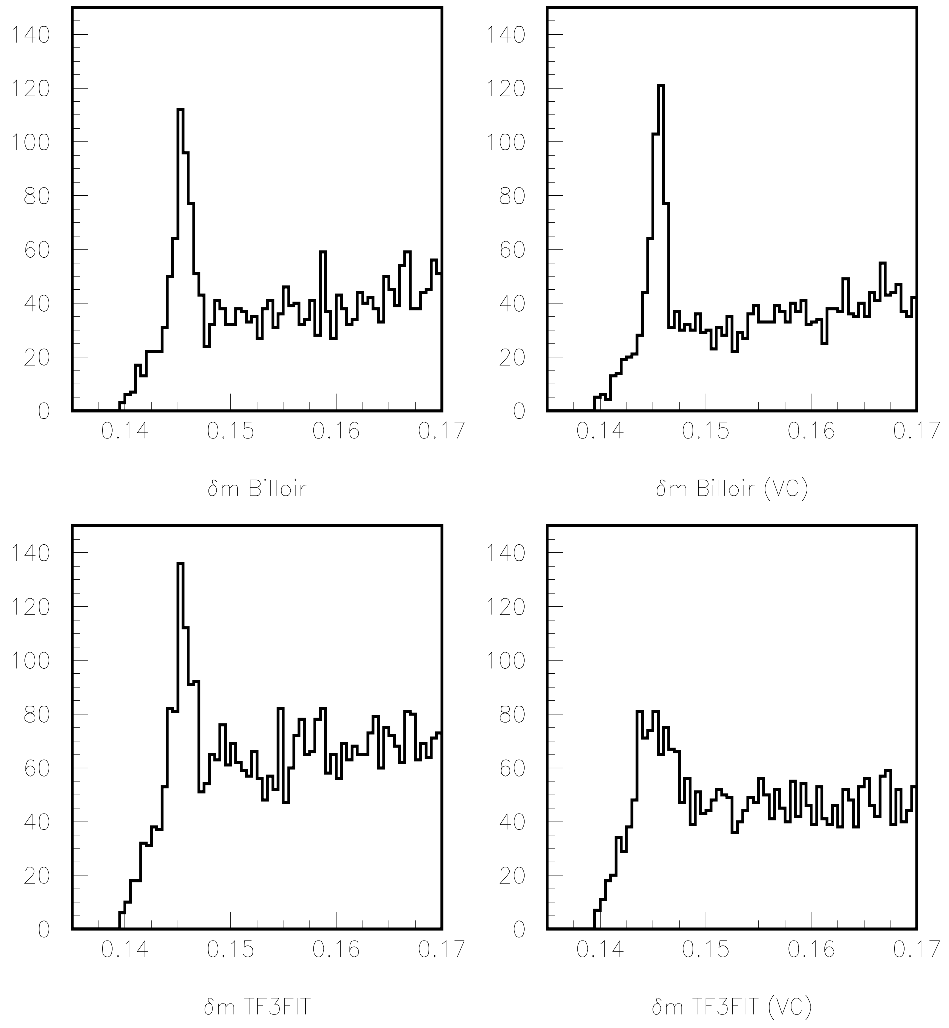

Figure 10.2: The $D^* - D$ mass difference in data. The top two plots use the Billoir fitter. The lower two plots show the same plots for the old track fitter, TF3FIT. The plots to the right also have a vertex constrained (VC) fit performed. The vertex constrained fit improves the resolution with the Billoir fitter, while the data from TF3FIT is worse when the vertex constrained fit is performed.

Appendix A

Track model

This section describes the conventions that are used in the description of the track in the Billoir fitter. The track model is, locally, a helix, described by five track parameters, $\eta = (\mathrm{K}, \phi, d, t, z)$. This discussion follows closely that of Ref. [49], but it is included here to make this document self contained.

K is the signed curvature ($= q/2\rho$), where ρ is the radius of curvature and q is the charge in units of the proton charge. Note that K is the "diameter of curvature". The sign of the curvature is the same as the charge of the particle. K is measured in m^{-1}.

ϕ is the azimuth of the momentum, P, at the point of closest approach of the track to the z-axis.

d is the signed 2-D distance of closest approach of the track to the z-axis. The sign convention is as follows. Complete the circle of the track in the $r - \phi$ ($x - y$) plane. If the origin is outside the circle then d and K have the same sign.

t is the tangent of the dip angle, $\tan\lambda$. Also often expressed as the cotangent of the polar angle, $\cot\theta$.

z is the value of z at the point of closet approach of the track to the z-axis.

In the Billoir fitter, it is very convenient to describe the track with respect to different coordinate systems along the track. For example it is very convenient to transport to the wire position when adding on an axial wire hit to the track. However, these other coordinate systems are only linear translations, not rotations, of the global coordinate system. This means that the coordinate axes in the local coordinate systems are parallel to the axes of the global coordinate system.

From the track parameters and the magnetic field strength the components of the particle's 3-momentum are calculated

$$P_x = \frac{A}{|\mathrm{K}|}\cos\phi, \tag{A.1}$$

$$P_y = \frac{A}{|\mathrm{K}|}\sin\phi, \tag{A.2}$$

$$P_z = \frac{A}{|\mathrm{K}|}t, \tag{A.3}$$

where $A = cB/2$. c is the speed of light and B is the magnetic field, the factor of 2 comes from the fact that K is the "diameter of curvature". A 1.5 Tesla magnetic field gives $A = 0.2248425$ GeV/mc. This is in units such that the momentum is expressed in GeV/c and K is in m^{-1}.

For many purposes it is useful to parametrize the helix in terms of the track parameters. This is used repeatedly throughout this work for calculation of derivatives, intercepts of tracks with detector elements, etc. The helix is most easily parametrized in terms of the turning angle, ψ, from the point of closest approach of the track to the z-axis

$$x_t(\psi) = -d\sin\phi + \frac{1}{2\mathrm{K}}(\sin(\phi + q\psi) - \sin\phi), \tag{A.4}$$

$$y_t(\psi) = d\cos\phi + \frac{1}{2\mathrm{K}}(\cos\phi - \cos(\phi + q\psi)), \tag{A.5}$$

$$z_t(\psi) = z + t\rho\psi = z + \frac{q\psi}{2\mathrm{K}}t. \tag{A.6}$$

This is generally a convenient form, however, as the curvature, K, goes to zero, ψ goes to zero for a constant arc length along the track. The equations are well behaved in this limit but they become numerically instable. This problem can be solved by parametrizing the track as a function of the arc length instead of the turning angle. Using this the track parametrization becomes numerically stable even as the curvature goes to zero, i.e., for a straight line. This is a nice feature, and is especially desirable in patter recognition. However, it is not a problem in the current implementation of the fitter, as all operations, like transport and evaluation of derivatives are done for a seed track which is well behaved.

It is sometimes convenient, e.g., for plotting, to have the center of the track circle in the xy-plane

$$
\begin{aligned}
x_c &= -(d + \frac{1}{2\mathrm{K}}) \sin\phi, && \text{(A.7)} \\
y_c &= \quad (d + \frac{1}{2\mathrm{K}}) \cos\phi. && \text{(A.8)}
\end{aligned}
$$

Appendix B

Transport derivatives

The derivatives needed for the transport are derived here. All derivatives are exact, no approximations are made that assumes that the transport distance is small compared to the radius of curvature. As discussed in a Section 9.5.1 a transport is described by a change of the reference point of a track, this change is $(\Delta x, \Delta y, \Delta z)$. The initial parameters are described by $\eta = (\mathrm{K}, \phi, d, t, z)$ and the final parameters are denoted by $\eta' = (\mathrm{K}', \phi', d', t', z')$. The first task is to find $\eta' = \eta'(\eta)$. This is a geometrical problem, see Figure 9.3. With the shorthand notation

$$q = \frac{\mathrm{K}}{|\mathrm{K}|}, \tag{B.1}$$

$$x'_c = \Delta x - \left(d + \frac{1}{2\mathrm{K}}\right) \sin\phi, \tag{B.2}$$

$$y'_c = \Delta y + \left(d + \frac{1}{2\mathrm{K}}\right) \cos\phi, \tag{B.3}$$

$$r'_c = \sqrt{x'^2_c + y'^2_c}, \tag{B.4}$$

where x'_c and y'_c are the center of the track circle in the $r - \phi$ plane in the new coordinate system. Using this the track parameters at the new reference point are given by

$$\mathrm{K}' = \mathrm{K}, \tag{B.5}$$

$$d' = q r'_c - \frac{1}{2\mathrm{K}}, \tag{B.6}$$

$$\phi' = \arctan\frac{x'_c}{y'_c}, \tag{B.7}$$

$$z' = z + \Delta z - \frac{t\Delta\phi}{2\mathrm{K}}, \tag{B.8}$$

$\frac{\partial \rightarrow}{\partial \downarrow}$	K'	ϕ'	d'	t'	z'
K	1	$-\frac{a}{2K^2 r_c'^2}$	$\left[1-\frac{qb}{r_c'}\right]\frac{1}{2K^2}$	0	$\frac{t}{2K}\left[\frac{\Delta\phi}{K}+\frac{\partial\phi'}{\partial K}\right]$
ϕ	0	$\left[1+\frac{1}{2K}\right]\frac{c}{r_c'^2}$	$-\left[1+\frac{1}{2K}\right]\frac{qa}{r_c'^2}$	0	$\frac{t}{2K}\left[\frac{\partial\phi'}{\partial\phi}-1\right]$
d	0	$\frac{a}{r_c'^2}$	$\frac{qc}{r_c'}$	0	$\frac{t}{2K}\frac{\partial\phi'}{\partial d}$
t	0	0	0	1	$-\frac{\Delta\phi}{2K}$
z	0	0	0	0	1

Table B.1: In the table the following shorthands are used: $a = y_c' \sin\phi + x_c' \cos\phi$, $b = y_c' \cos\phi + x_c' \sin\phi$, and $c = y_c' \cos\phi - x_c' \sin\phi$.

$$t' = t, \tag{B.9}$$

where $\Delta\phi = \phi - \phi'$. It is now straight forward to evaluate the needed derivatives and the results are shown in Table B.1.

Bibliography

[1] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963); M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[2] Phys. Rev. D 50, 1315 (1994) and references therein.

[3] M. Neubert, Phys. Lett. **B264**, 455 (1991).

[4] H. Albrecht *et al.*, (ARGUS), Nucl. Inst. Meth. **A275**, 1 (1989).

[5] Y. Kubota *et al.*, (CLEO), Nucl. Instr. and Meth. **A320**, 66 (1992).

[6] B. Barish *et al.*, (CLEO), Phys. Rev. Lett. **76**, 1570 (1996).

[7] G. Altarrelli *et al.*, Nucl. Phys. **B208**, 365 (1982).

[8] CLEO Collaboration, CLEO-CONF-95-10, EPS0166 (1995).

[9] B. Barish *et al.*, (CLEO), Phys. Rev. **D51**, 1014 (1995).

[10] D. Buskulic *et al.*, (ALEPH), Phys. Lett. **B345**, 103 (1995).

[11] R. Akers *et al.*, (OPAL), "A study of Charm Meson Production in Semileptonic B Decays", CERN Preprint No. CERN-PPE/95-02 (1995) (unpublished).

[12] See P. Renton, *Electroweak Interactions*, Cambridge University, New York (1990).

[13] S.P. Booth, *et al.* (UKQCD), Phys. Rev. Lett. **72**, 462 (1994); C. Bernard, Y. Shen, and A. Soni, Phys. Lett. B **317**, 164 (1993); B. Blok and M. Shifman, Phys. Rev. D **47**, 2949 (1993); other predictions are summarized in Ref. [50].

[14] N. Isgur and M.B. Wise, Phys. Lett. **B232**, 113 (1989).

[15] N. Isgur and M.B. Wise, Phys. Lett. **B237**, 527 (1990).

[16] S. Sanghera *et al.*, (CLEO), Phys. Rev. **D47**, 791 (1993).

[17] H. Albrecht *et al.*, (ARGUS), Z. Phys. **C57**, 533 (1993).

[18] N. Isgur, D. Scora, B. Grinstein, and M.B. Wise, Phys. Rev. **D39**, 799 (1989).

[19] J.G. Körner and G.A. Schuler, Z. Phys. **C38**, 511 (1988).

[20] F.J. Gilman and R.L. Singleton, Phys. Rev. **D41**, 142 (1990).

[21] J.D. Richman, Cal. Tech. Preprint CALT-68-1148, (1984) (unpublished).

[22] M. Wirbel *et al.*, Z. Phys. **C29**, 637 (1985).

[23] D. Scora and N. Isgur, Phys. Rev. **D52**, 2783 (1995).

[24] M. Neubert, Phys. Rep. **245**, 259 (1993).

[25] F.E. Close and A. Wambach, Nucl. Phys. **B412**, 169 (1994).

[26] M.E. Luke, Phys. Lett. **B252**, 447 (1990).

[27] Hadronic events are identified as having atleast three charged tracks and 0.15 $E_{\rm cm}$ energy deposited in the calorimeter.

[28] J.C. Anjos *et al.*, (E691), Phys. Rev. **D48**, 56 (1993).

[29] G.C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978).

[30] TRKSIM is a fast Monte Carlo that uses measured efficiencies and resolutions to generate parametrized detector response for events.

[31] S.L. Jones, "Semileptonic B Meson to Excited Charm Meson Decay", Ph. D. Thesis, Cornell University (1996) (unpublished).

[32] P. Avery *et al.*, (CLEO), Phys. Lett. **B331**, 236 (1994).

[33] D.M. Schmidt, "Measurement of Charm Semileptonic Form Factors", Ph. D. Thesis, University of California Santa Barbara (1992) (unpublished).

[34] D.M. Schmidt, R.M. Morrison, and M. Witherell, Nucl. Instr. and Meth. **A328**, 547 (1993).

[35] See W.T. Eadie *et al.*, *Statistical Methods in Experimental Physics*, Elsevier, New York (1988).

[36] GEANT Detector Description and Simulation Tool, Version 3.15, CERN Program Library W5103, CERN (1993) (unpublished).

[37] D. Lange and A. Ryd, CLEO Internal Note, CBX 96-6 CSN 95-345 (1996) (unpublished).

[38] S. Menary, CLEO Internal Note, CBX 92-103 (1992) (unpublished).

[39] M. Neubert and V. Rieckert, Nucl. Phys. **B382**, 97 (1992).

[40] P. Avery, CLEO Internal Note, CBX 91-74 (1991) (unpublished).

[41] P. Avery, CLEO Internal Note, CBX 92-40 (1992) (unpublished).

[42] P. Billoir, Nucl. Instr. Meth. **225**, 352 (1983).

[43] P. Billoir and S. Qian, Nucl. Instr. Meth. **A225**, 219 (1990).

[44] R.E. Kalman, J. Basic Eng. **82**, 34 (1961); R.E. Kalman and R.S. Bucy, ibid., 95 (1961).

[45] R. Fruhwirth, HEPHY-PUB-516/88, "Application of Filter Methods to the Reconstruction of Tracks and Vertices in Events of Experimental High Energy Physics", Ph. D. Thesis (1988) (unpublished).

[46] Phys. Rev. D 50, 1253 (1994).

[47] R. Kutschke and A.Ryd, CLEO Internal Note, CBX 96-20 (1996) (unpublished).

[48] R. Kutschke and A. Ryd, CLEO Internal Note, CBX 96-XYZ (1996) (unpublished).

[49] R. Kutschke, CLEO Internal Note, CSN 94-334 (1992) (unpublished).

[50] J.D. Richman and P. Burchat, Rev. Mod. Phys. **67**, 893 (1995).