A Measurement of the Branching Fractions and CP Asymmetries of the Decays $B^{\pm} \to K^{*\pm}\pi^{0}$ and $B^{\pm} \to \rho^{\pm}\pi^{0}$ at *BABAR*

James Robert Gaillard

Imperial College London

A thesis to be submitted for the degree of
Doctor of Philosophy
of The University of London
and the Diploma of Imperial College.

Abstract

Measurements of the branching fractions (averaged over charge-conjugate decays), $\mathcal{B}$, and CP violating charge asymmetries, $\mathcal{A}_{CP}$, of the decays $B^{\pm} \to K^{*\pm}\pi^{0}$ and $B^{\pm} \to \rho^{\pm}\pi^{0}$ are presented :

$$
\begin{aligned}
\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^{0}) &= [6.9\ \pm 2.1\ (Stat.)\ \pm 1.2\ (Syst.)] \times 10^{-6} \\
\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^{0}) &= 0.03\ \pm 0.29\ (Stat.)\ \pm 0.05\ (Syst.)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{B}(B^{\pm} \to \rho^{\pm}\pi^{0}) &= [10.3\ \pm 1.4\ (Stat.)\ \pm 0.9\ (Syst.)]\ \times 10^{-6} \\
\mathcal{A}_{CP}(B^{\pm} \to \rho^{\pm}\pi^{0}) &= 0.03\ \pm 0.13\ (Stat.)\ \pm 0.02\ (Syst.)
\end{aligned}
$$

The B decays are reconstructed from the three body final states $K^{\pm}\pi^{0}\pi^{0}$ and $\pi^{\pm}\pi^{0}\pi^{0}$ respectively. The analysis uses approximately 232 million charged B decays produced by the PEP-II storage rings at SLAC and recorded by the *BABAR* detector over the period 1999-2004.

Acknowledgements

I owe many people a great deal of gratitude for their contributions to this thesis, and my life, over the past three and a half years.

Firstly I would like to give huge thanks to my supervisor, Paul Dauncey, for his interest, encouragement and wisdom.

I owe so much to Dan Bowerman in particular for his constant help and advice, and friendship over the past three years.

Geoff Morton, for all his efforts and help on the analysis.

Ivo Eschrich, for his input on the analysis and especially for his help during my time on the EMC. Rocky Eschrich for his zen.

Jordan Nash, for his supervision in the first year of my PhD.

Andreas Hoecker, Fergus Wilson and the three-body charmless AWG.

Paul, Julia Sedgbeer, Peter Dornan and PPARC for giving me the opportunity to study at Imperial.

My family in California, Gwen and Don, for their kindness and generosity during my time in the States.

The friends I made out in America and in London - Steve, Ed, Mike, Wah, Sian, Kelly, Tom, Mark, Gary, Matt Ryan, Chris, Lisa, Guy and Amber.

Rob Flack, who deserves his own space for never failing to find one for me to park in, and for his tireless pessimism at five in the morning in the Britannia Arms, Cupertino.

Matt Noy, for all the swift halves and ones for the road.

My mum, dad and brother, for much more than I can give space to here.

Lastly, and most importantly, I would like to thank Vicky, whose love, compassion and patience made this possible.

Contents

Abstract 2

Acknowledgements 3

Contents 5

List of Figures 10

List of Tables 13

Chapter 1. Theoretical background 15

1.1 Introduction 15

1.2 The C, P and T operations 16

1.3 CP violation in the Standard Model 17

1.3.1 Glashow-Weinberg-Salam Model 18

1.3.2 CKM matrix 20

1.4 CP violation in the b-sector 22

1.4.1 Direct CP violation 24

1.4.2 CP violation in mixing - ‘indirect’ CP violation 25

1.4.3 CP violation in the interference of decays with and without mixing 27

1.4.4 Measuring β 29

1.4.5 Measuring α 29

1.4.6 Measuring γ 31

1.4.7 State of the art 32

1.5 Charmless B physics 32

1.5.1 Approximate theories 34

1.5.2 Long distance rescattering effects for $K\pi$ channels 36
1.6 Theoretical predictions for $B^{\pm} \to K^{*\pm}\pi^0$ and $B^{\pm} \to \rho^{\pm}\pi^0$ 36
1.7 Kinematics of three-body decays 37
1.8 Resonance kinematics, amplitudes and interference 38
1.8.1 Quasi two-body analysis 41

Chapter 2. The PEP-II Storage Rings and the BABAR Detector 42
2.1 Introduction 42
2.2 The PEP-II Storage Rings 43
2.2.1 Beam parameters 44
2.2.2 Interaction region 44
2.3 The *BABAR* detector 45
2.4 Silicon Vertex Tracker (SVT) 46
2.4.1 Design 48
2.4.2 Performance and calibration 49
2.5 Drift Chamber (DCH) 50
2.5.1 Design 51
2.5.2 Performance and calibration 53
2.6 Detector of Internally Reflecting Cherenkov Light (DIRC) 53
2.6.1 Design 54
2.6.2 Performance and calibration 56
2.7 Electromagnetic Calorimeter (EMC) 57
2.7.1 Design 57
2.7.2 Performance and calibration 60
2.8 Instrumented Flux Return (IFR) 62
2.8.1 Design 62
2.8.2 Performance 63
2.9 Trigger 64
2.9.1 L1 trigger 65
2.9.2 L3 trigger 65

Chapter 3. Event selection and continuum background fighting 67

3.1 Introduction 67

3.2 Data and Monte Carlo samples 68

3.3 Event selection 69

3.3.1 B meson kinematics at *BABAR* 69

3.3.2 Event topology of B decays 72

3.3.3 Event preselection 75

3.4 Candidate reconstruction 76

3.4.1 π^0 reconstruction 76

3.4.2 Track identification 79

3.4.3 K^* and ρ mass selection 81

3.5 Continuum background fighting 86

3.5.1 Introduction 86

3.5.2 Linear discriminant 87

3.5.3 Artificial neural network 88

3.5.4 Selecting the best input variables 90

3.5.5 Correlation of variables with ΔE, m_{ES}. 94

3.5.6 MVA designs 96

3.5.7 MVA sample composition 97

3.5.8 Optimal ANN configuration 98

3.5.9 Comparison of ANN and LDA 99

3.5.10 MVA conclusions 101

3.6 Multiple candidates 102

3.6.1 Final algorithm 106

3.7 Summary of cuts and efficiencies 108

Chapter 4. Analysis method and validation 112

4.1 Introduction 112

4.2 The maximum likelihood fit 112

4.2.1 MINUIT and RooFit 113

4.3 Fit implementation 114

4.3.1 The event PDF and likelihood 114

4.3.2 Description of sub-component PDFs 117
4.4 PDF Projections 118
4.5 B–related background 118
4.5.1 Higher $K^{**\pm}$, $\rho^{*\pm}$ and non-resonant backgrounds 124
4.5.2 Effect of possible interfering amplitudes on the branching fraction measurements from non–resonant decays 125
4.5.3 Estimation of higher resonance pollution 128
4.5.4 Summary of B backgrounds 131
4.6 Fit validation 134
4.6.1 Fits to individual modes 134
4.6.2 Toy Monte Carlo studies 134
4.6.3 Toy studies using full Monte Carlo simulation 136
4.6.4 Optimisation of ANN 139
4.7 Blind fits to data 142

Chapter 5. Results and conclusions 144
5.1 Introduction 144
5.2 Analysis of systematic uncertainties 144
5.2.1 Systematic uncertainties on the yields 145
5.2.2 Systematic uncertainties on the branching fraction determination 150
5.2.3 Systematic uncertainties on $\mathcal{A}_{CP}$ 153
5.3 Results from fit 157
5.3.1 Fit results for $B^{\pm} \to K^{*\pm}\pi^0$ 157
5.3.2 Fit results for $B^{\pm} \to \rho^{\pm}\pi^0$ 159
5.3.3 Stability cross-check 161
5.4 Final results inclusive of systematics 161
5.4.1 Signal–enhanced discriminating variable projections 163
5.5 Conclusions 168

References 173

List of Figures

1.1 The unitarity triangle relevant to the decays of B mesons. 23
1.2 Illustration of the need for differing strong and weak phases for direct CP violation. 25
1.3 $B^0 - \overline{B^0}$ mixing. 26
1.4 Tree and penguin diagrams for $\sin 2\beta$ 29
1.5 Tree and penguin diagrams for processes used in the measurement of $\sin 2\alpha$ 30
1.6 World average confidence levels of the unitarity triangle in the $\overline{\rho}, \overline{\eta}$ plane. 33
1.7 Feynman diagrams describing charmless B decays. 34
1.8 A generic Dalitz plot. 39

2.1 A schematic of the PEP-II elements at the *BABAR* IP. 46
2.2 Cross sectional projections of the *BABAR* detector. 47
2.3 Schematic of SVT. 48
2.4 Projection along z of the SVT silicon layer arrangement. 49
2.5 SVT hit resolutions vs z and ϕ. 50
2.6 Schematic of DCH. All dimensions are in mm. 51
2.7 Arrangement of DCH superlayers. 52
2.8 Transverse momentum resolution of tracks in the DCH. 53
2.9 $\mathrm{d}E/\mathrm{d}x$ as a function of momentum in the DCH. 54
2.10 DCH efficiency as a function of polar angle and p_t. 55
2.11 Illustration of the DIRC principle. 56
2.12 DIRC timing cut plots. 57
2.13 Kaon-pion separation in the DIRC as a function of momentum. 58
2.14 Kaon efficiency and pion rejection as a function of momentum. 58

2.15 Cross section of EMC. 59

2.16 Cut-away of EMC crystal layout. 60

2.17 The energy resolution (a) and the angular resolution (b) of the EMC as a function of energy. 61

2.18 Illustration of IFR assembly. All dimensions are in mm. 63

2.19 The muon efficiency and pion misidentification rates of the IFR as a function of polar angle (a) and momentum (b). 64

3.1 Distribution of MC events in the (ΔE,m_{ES}) plane for correctly reconstructed $B^{\pm} \to \rho^{\pm}\pi^{0}$ signal events. 71

3.2 Distribution of MC events in the (ΔE,m_{ES}) plane for $B^{0} \to \rho^{\pm}\rho^{\mp}$ (left) and $B^{\pm} \to \pi^{\pm}\pi^{0}$ (right) events reconstructed as $B^{\pm} \to \rho^{\pm}\pi^{0}$. 72

3.3 Distribution of $|\cos\theta^{B}_{Sph}|$ for signal $B^{\pm} \to K^{*\pm}\pi^{0}$ MC events and continuum MC. 74

3.4 Distribution of π^{0} invariant mass for non-resonant $\pi^{\pm}\pi^{0}\pi^{0}$ simulated events (left) and for data events (right) which pass the preselection. 78

3.5 Data-MC comparison for the widths and means which are used in the γ energy smearing and shifting corrections. 79

3.6 Performance of kaon 'Loose' selection as a function of momentum for positive tracks (left) and negative tracks (middle). The overall data:MC comparison is shown on the right. 81

3.7 $K^{\pm}\pi^{0}$ invariant mass for truth-matched signal. Overlaid is a fitted Breit-Wigner lineshape, which describes the distribution well. 82

3.8 Distribution of invariant masses for the two $K^{\pm}\pi^{0}$ pairs (a) and $\pi^{\pm}\pi^{0}$ pairs (b) in the event. 85

3.9 Architecture of a single-output, single hidden-layer feed-forward neural network. 89

3.10 Discriminating variables for signal. 95

3.11 Performance plot of various ANN configurations showing continuum selection efficiency against signal selection efficiency for cuts applied to the output from the ANN. 98

3.12 The ANN output for signal and continuum events 99

3.13 Examples of the output for signal and background for the two linear discriminants. 100
3.14 Comparison of discriminating power 101
3.15 Plot of the ρ $\cos\theta_{Hel}$ for signal $\rho^{\pm}\pi^{0}$ correct-charge SCF and wrong-charge SCF. 105

4.1 PDF projections for the m_{ES} distributions. 119
4.2 PDF projections for the ΔE distributions. 120
4.3 PDF projections for the ANN_{Tran} distributions. 121
4.4 PDF projections for the $\pi^{\pm}\pi^{0}$ lineshape. 122
4.5 Dalitz plot showing K^{*} region and region from which the number of non–resonant decays is estimated. 126
4.6 Dalitz plot showing $\rho^{\pm}$ region and region from which the number of non–resonant decays is estimated. 127
4.7 $B^{\pm} \to K^{*\pm}\pi^{0}$ pull distributions from toy MC studies. 137
4.8 $B^{\pm} \to \rho^{\pm}\pi^{0}$ pull distributions from toy MC studies. 138
4.9 Pull distributions from toy MC studies using full MC for the $B^{\pm} \to K^{*\pm}\pi^{0}$ signal mode. 139
4.10 Pull distributions from toy MC studies using full MC for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ signal mode. 140
4.11 Fit to the distribution of significance. 141
4.12 The mean of the significance distributions for different cuts on ANN. 141
4.13 Distribution of likelihoods from toy MC experiments seeded with values from a blind fit to the data. The top plot is for the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis, the bottom is for $B^{\pm} \to \rho^{\pm}\pi^{0}$. 143

5.1 Likelihood curves for N^{Signal} and A^{Signal} for the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis. 163
5.2 Likelihood curves for N^{Signal} and A^{Signal} for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis. 164
5.3 Plots of the fit variables with a cut on likelihood ratio for the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis. 166
5.4 Plots of the fit variables with a cut on likelihood ratio for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis. 167

List of Tables

1.1 Isospin decomposition of $B \to \rho\pi$ in terms of the isospin amplitudes $A_{\Delta I, I_f}$. 31

2.1 Performance summary of the PEP-II machine. 44

2.2 Performance summary of the SVT. 49

3.1 Summary of π^0 selection requirements. The composite π^0 masses are fitted assuming production at the primary vertex. 77

3.2 Summary of track preselection requirements. 80

3.3 Summary of ρ mass cut investigations. 84

3.4 Values of the variables' discrimination power for truth-matched signal and non-resonant modes. 93

3.5 Values of the variables' signal discrimination power for the truth-matched signal and non-resonant modes. 93

3.6 Values of the variable correlations with ΔE, m_{ES} for the truth-matched $K^{\pm}\pi^0\pi^0$. 96

3.7 Values of the variable correlations with ΔE, m_{ES} for the truth-matched $K^{*\pm}\pi^0$. 96

3.8 Summary of best candidate strategies. 106

3.9 The efficiencies of the selection cuts for signal modes. 109

3.10 The efficiencies of the selection cuts for the data samples 110

3.11 The efficiencies of the selection for MC simulated background samples 111

4.1 $K^{**\pm}$ resonances decaying to $K\pi$ final states 128

4.2 Summary of the effect of the choice of $K^{**\pm}$ mass range. 130

4.3 Summary of the effect of the choice of ρ^* mass range. 131

4.4 Summary of B backgrounds to $B^{\pm} \to K^{*\pm}\pi^{0}$. 132

4.5 Summary of B backgrounds to $B^{\pm} \to \rho^{\pm}\pi^{0}$. 133

4.6 Results of the full fit applied to various MC samples 135

4.7 Results of the full fit applied to a single MC samples composed of generic $B(b \to c)$, continuum and signal. 135

5.1 Table of significant contributions from background normalisations to the systematic uncertainty on the $B^{\pm} \to K^{*\pm}\pi^{0}$ signal yield. 146

5.2 Table of significant contributions from background normalisations to the systematic uncertainty on the $B^{\pm} \to \rho^{\pm}\pi^{0}$ signal yield. 146

5.3 Table of systematic errors on the yields. 151

5.4 Table of systematic errors on the branching fraction determination. 154

5.5 Table of systematic errors for the asymmetry. 157

5.6 Summary of fit results for $B^{\pm} \to K^{*\pm}\pi^{0}$. 159

5.7 Summary of fit results for $B^{\pm} \to \rho^{\pm}\pi^{0}$. 161

Chapter 1

Theoretical background

1.1 Introduction

There exist three discrete transformations of fundamental importance in our understanding of physics at the smallest scales. They are the parity, charge-conjugation and time-reversal transformations, represented respectively in quantum mechanics by the two unitary operators P and C, and the antiunitary operator T. Classically the P and T space-time operations correspond to (P : $\vec{\mathbf{x}} \rightarrow -\vec{\mathbf{x}}$) and (T : $t \rightarrow -t$). The charge-conjugation transformation transposes particle for antiparticle, in essence changing the sign of all internal quantum numbers and phases. Although there is no strict classical representation of the charge-conjugation transformation (since it emerges from relativistic quantum mechanics), the transformation nevertheless has classical consequences. In electromagnetism, electric charges change sign under the transformation, reversing the direction of their associated currents and hence their electromagnetic fields.

Examples of fundamental interactions can be found in nature which are not symmetric under parity and charge-conjugation transformations separately [1]. In these instances it is said the C and P symmetries are 'violated'. In addition it has been known since 1964 that the product CP is violated in the decays of neutral kaons [2].

More recently the *BABAR* [3] and *BELLE* [4] experiments have succeeded in being the first experiments to claim discovery of CP violation in the b quark sector [5] [6], helping us to further understand the mechanism by which CP violation is manifested. Our current understanding of CP violation is incorporated into the Standard Model (SM) by the presence of the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [7] in the electroweak Lagrangian [8] [9] [10].

Besides the weak decays of kaons and B mesons, the preponderance of matter over antimatter in the visible universe is a strong indication that at some level CP was not a good symmetry of the early evolving universe [11].

Despite the fact that these symmetries do not always hold, it is understood that any system that can be described by a local quantum field theory that preserves causality must have an exact symmetry under the transformation represented by the product of operations CPT [12]. This implies any system that violates CP must also violate T.

In this chapter the mechanism for CP violation in the Standard Model is discussed before setting it in the context of the decays of B mesons. After this discussion, additional physics relating to the specific decays under consideration in this thesis are introduced.

1.2 The C, P and T operations

In discussing how elementary particles and their interactions behave under the transformations P, C and T it is appropriate to consider how the operations are represented in our quantum description of nature.

Drawing analogies from classical expectation, one is able to postulate the following transformation properties:

$$PA^{\mu}(t,\vec{\mathbf{x}})P^{\dagger} = A_{\mu}(t,-\vec{\mathbf{x}}); \quad PJ^{\mu}(t,\vec{\mathbf{x}})P^{\dagger} = J_{\mu}(t,-\vec{\mathbf{x}}); \quad P\partial^{\mu}P^{\dagger} = \partial_{\mu} \tag{1.1}$$

$$CA^{\mu}(t,\vec{\mathbf{x}})C^{\dagger} = -A^{\mu}(t,\vec{\mathbf{x}}); \quad CJ^{\mu}(t,\vec{\mathbf{x}})C^{\dagger} = -J^{\mu}(t,\vec{\mathbf{x}}); \quad C\partial^{\mu}C^{\dagger} = \partial^{\mu} \tag{1.2}$$

$$TA^{\mu}(t,\vec{\mathbf{x}})T^{\dagger} = A_{\mu}(-t,\vec{\mathbf{x}}); \quad TJ^{\mu}(t,\vec{\mathbf{x}})T^{\dagger} = J_{\mu}(-t,\vec{\mathbf{x}}); \quad T\partial^{\mu}T^{\dagger} = -\partial_{\mu}, \tag{1.3}$$

where A^{μ} represents a spin-1 vector gauge field, J^{μ} represents a general field current and $\partial^{\mu} = \partial/\partial x_{\mu}$. Using the transformation ansatz for the current J^{μ}, one may derive the transformation properties of the fields contributing to the currents. For instance, in the case that $J^{\mu} = \overline{\psi}(t,\vec{\mathbf{x}})\gamma^{\mu}\psi(t,\vec{\mathbf{x}})$ represents the current for a Dirac field $\psi(t,\vec{\mathbf{x}})$, it can be shown [13] that the field must transform as

$$P\psi(t,\vec{\mathbf{x}})P^{\dagger} = \gamma_0\psi(t,-\vec{\mathbf{x}}); \quad P\overline{\psi}(t,\vec{\mathbf{x}})P^{\dagger} = \overline{\psi}(t,-\vec{\mathbf{x}})\gamma_0; \tag{1.4}$$

$$C\psi(t,\vec{\mathbf{x}})C^{\dagger} = i\gamma^2\gamma^0\overline{\psi}^T(t,\vec{\mathbf{x}}); \quad C\overline{\psi}(t,\vec{\mathbf{x}})C^{\dagger} = -i\overline{\psi}^*(t,\vec{\mathbf{x}})\gamma^2; \tag{1.5}$$

$$T\psi(t,\vec{\mathbf{x}})T^{\dagger} = \gamma^1\gamma^3\psi(-t,\vec{\mathbf{x}}); \quad T\overline{\psi}(t,\vec{\mathbf{x}})T^{\dagger} = \overline{\psi}(-t,\vec{\mathbf{x}})\gamma^1\gamma^3; \tag{1.6}$$

to satisfy the current terms in Eq.(1.1) - Eq.(1.3), where the γ^{μ} matrices are those of the Dirac-Pauli representation.

It is relatively straightforward to derive similar relations for spin-0 fields and other bilinear quantities (though of course this method is not the only way to derive them). Equipped with such transformations one is then able to tackle the issue of whether a given Lagrangian is invariant under the C, P and T operations or any combination thereof. A Lagrangian of particular interest for such a question is that of the Standard Model.

1.3 *CP* violation in the Standard Model

The Standard Model of the strong and electroweak forces is described by the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where the subscripts denote colour, left-handedness and weak hypercharge respectively. The SM Lagrangian can be written as

$$\mathcal{L}_{SM} = \mathcal{L}_{SU(3)_C} + \mathcal{L}_{SU(2)_L \times U(1)_Y}, \tag{1.7}$$

where $\mathcal{L}_{SU(3)_C}$ and $\mathcal{L}_{SU(2)_L \times U(1)_Y}$ are the QCD and electroweak Lagrangians respectively. In writing the most general Lagrangian for QCD, one must include a gauge field interaction term $G\widetilde{G}$ which violates both parity and time reversal invariance. This contribution is usually set to zero 'by hand' thereby imposing parity and *CP* conservation, though this is generally regarded as unnatural. The presence (or rather absence) of this term is known as the 'Strong *CP*' problem.

Leaving QCD aside, we focus now on the issue of *CP* violation in the electroweak Lagrangian which is the only source of *CP* violation in the SM.

1.3.1 Glashow-Weinberg-Salam Model

The full electroweak Lagrangian of the Glashow-Weinberg-Salam (GWS) Model [8] [9] [10] can be written as

$$
\begin{aligned}
\mathcal{L}_{GWS} &= -\frac{1}{4}\mathbf{W}_{\mu\nu} \cdot \mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\
&+ \overline{\chi_L}\gamma^\mu(i\partial_\mu - g\frac{1}{2}\boldsymbol{\tau} \cdot \mathbf{W}_\mu - g'\frac{Y}{2}B_\mu)\chi_L \\
&+ \overline{q_R}\gamma^\mu(i\partial_\mu - g'\frac{Y}{2}B_\mu)q_R \\
&+ |(i\partial_\mu - g\frac{1}{2}\boldsymbol{\tau} \cdot \mathbf{W}_\mu - g'\frac{Y}{2}B_\mu)\phi|^2 - V(\phi) \\
&- \sum_{i,j=1}^{3}(\Gamma^u_{ij}\overline{\chi_i}_L\widetilde{\phi}u_{j_R} + \Gamma^d_{ij}\overline{\chi_i}_L\phi d_{j_R} + \Gamma^l_{ij}\overline{\chi_i}_L\phi l_{j_R} + h.c.).
\end{aligned}
\tag{1.8}
$$

Here $\boldsymbol{W}^\mu = (W^{\mu 1}, W^{\mu 2}, W^{\mu 3})$ and B^μ define the isotriplet $SU(2)$ and singlet $U(1)$ gauge fields of the model respectively, with $\mathbf{W}_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \boldsymbol{W}_\mu - g\boldsymbol{W}_\mu \times \boldsymbol{W}_\nu$ and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. $\boldsymbol{\tau}$ and Y are the generators of the $SU(2)$ and $U(1)$ groups, while g and g' are the respective gauge couplings. Fermion fields couple to the $SU(2)$ and $U(1)$ gauge fields as left-handed doublets represented by $\chi_{i_L} = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L$ for the quarks and $\begin{pmatrix} \nu_i \\ l_i \end{pmatrix}_L$ for leptons with i denoting the fermion generation, of which there are three in the SM. Additionally the fermions couple to the $U(1)$ field as right-handed singlets $q_{i_R} = u_{i_R}, d_{i_R}, l_{i_R}$; notice there is no right-handed neutrino in the SM. The GWS model contains a single $SU(2)$ doublet ϕ of Higgs scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad \text{with} \quad \widetilde{\phi} = \begin{pmatrix} \phi^0 \\ -\phi^- \end{pmatrix} \tag{1.9}$$

The shape of the scalar potential $V(\phi)$ is such that the Higgs fields acquire vacuum expectation values which we 'choose' to be $\langle 0|\phi^0|0\rangle = v + h(x)$ and $\langle 0|\phi^+|0\rangle = \langle 0|\phi^-|0\rangle = 0$, thereby breaking the $SU(2)$ and $U(1)_Y$ gauge symmetries. However, this choice preserves the symmetry $\tau^3 + Y/2 = Q$ thereby conserving charge and keeping the photon massless. The symmetries $(\tau_1 \pm i\tau_2)$ and $(Y/2 - \tau^3)$ are broken, giving rise to the three massive gauge bosons $(W^\pm, Z^0)$.

The non-zero vacuum expectation value for the neutral Higgs field also gives masses to the fermions by way of the last term in Eq.(1.8) (known as the Yukawa coupling term) which becomes

$$(\Gamma^u_{ij}\overline{\chi_i}_L\widetilde{\phi}u_{j_R} + \Gamma^d_{ij}\overline{\chi_i}_L\phi d_{j_R} + h.c.) \quad \rightarrow \quad (M^u_{ij}\overline{u_i}_L u_{j_R} + M^d_{ij}\overline{d_i}_L d_{j_R} + h.c.) \tag{1.10}$$

where

$$M^u_{ij} = v\Gamma^u_{ij} \quad \text{and} \quad M^d_{ij} = v\Gamma^d_{ij} \tag{1.11}$$

are the mass matrices, which are in general complex. In their current (arbitrary) form, the mass matrices are not particularly useful. However if we diagonalise them such that the elements are real, the diagonal elements will be the fermion masses and the eigenvectors of the diagonalised matrix will be the fermion mass eigenstates. We do this with the help of four unitary matrices, T^u_L, T^u_R, T^d_L and T^d_R.

$$T^u_L M^u T^{u\dagger}_R = M^u_{diag}; \quad T^d_L M^d T^{d\dagger}_R = M^d_{diag} \tag{1.12}$$

with eigenstates

$$u^m_{L(R)} = T^u_{L(R)}u_{L(R)}; \quad d^m_{L(R)} = T^d_{L(R)}d_{L(R)}. \tag{1.13}$$

$$\tag{1.14}$$

These can be inverted to give

$$u_{L(R)} = T^{u\dagger}_{L(R)} u^m_{L(R)}; \quad d_{L(R)} = T^{d\dagger}_{L(R)} d^m_{L(R)}. \tag{1.15}$$

Substituting these expressions back into the Lagrangian which now has a broken symmetry, the interaction terms of the left-handed currents with the charged electroweak gauge fields are

$$\mathcal{L}_{int} = -\frac{g}{\sqrt{2}} \overline{u^m_i}_L T^u_{ik_L} \gamma^\mu W^+_\mu T^{d\dagger}_{kj_L} d^m_{jL} - \frac{g}{\sqrt{2}} \overline{d^m_i}_L T^d_{ik_L} \gamma^\mu W^-_\mu T^{u\dagger}_{kj_L} u^m_{jL} \tag{1.16}$$

where $W^\pm_\mu = W^1_\mu \pm iW^2_\mu$ of Eq.(1.8). The matrix product $T^u_{ik_L} T^{d\dagger}_{kj_L}$ is written as $\mathbf{V}_{ij_{CKM}}$, and is given the name the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

We can now apply our transformation rules to this interaction. Under *CP*, $\mathcal{L}_{int}$ transforms as

$$(CP)\mathcal{L}_{int}(P^\dagger C^\dagger) = -\frac{g}{\sqrt{2}} \overline{d^m_j}_L \gamma^\mu \mathbf{V}_{ij_{CKM}} W^-_\mu u^m_{iL} - \frac{g}{\sqrt{2}} \overline{u^m_{jL}} \gamma^\mu \mathbf{V}^*_{ij_{CKM}} W^+_\mu d^m_{iL} \tag{1.17}$$

Notice that all terms are interchanged except for the $\mathbf{V}_{ij_{CKM}}$ and $\mathbf{V}^*_{ij_{CKM}}$ terms. These are of course unaffected due to the *unitary* nature of the operators[1]. Thus the interaction Lagrangian is only guaranteed to be invariant under *CP* if the CKM matrix is real. As we shall see, for a unitary 3×3 matrix this is in general certainly not the case. The presence of an *irremovable* complex phase in $\mathbf{V}_{CKM}$ guarantees *CP* violation in the SM.

1.3.2 CKM matrix

It can be shown that a general 3×3 unitary matrix can be expressed in terms of three real parameters and six pure phases [14]. However, in the case we are considering we are free to rotate each quark field by an arbitrary global $U(1)$ phase without

[1] Recall a property of a unitary operator U is $U(a\phi) = a(U\phi)$ where a is a complex coefficient.

affecting any of the physics. This amounts to transforming $\mathbf{V}_{CKM}$ as

$$\mathbf{V}_{CKM} \rightarrow \begin{pmatrix} e^{-i\phi_1^u} & 0 & 0 \\ 0 & e^{-i\phi_2^u} & 0 \\ 0 & 0 & e^{-i\phi_3^u} \end{pmatrix} \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} e^{i\phi_1^d} & 0 & 0 \\ 0 & e^{i\phi_2^d} & 0 \\ 0 & 0 & e^{i\phi_3^d} \end{pmatrix} \tag{1.18}$$

Thus it would seem we have six phases to play with to cancel the six phases of the CKM matrix. However as we can always factorise any one of these phases, only five relative phases of $\mathbf{V}_{CKM}$ can actually be removed in this way. Thus $\mathbf{V}_{CKM}$ can be expressed completely in terms of three real parameters and one remaining arbitrary phase. This conclusion was first presented by Kobayashi and Maskawa in their paper [7] of 1973.

Having obtained the CKM matrix, it is a question for phenomenology to parameterise it in an appropriate way. Currently the magnitudes of the elements of the CKM matrix are (to a 90% confidence interval) [15]

$$\begin{pmatrix} 0.9739 - 0.9751 & 0.221 - 0.227 & 0.0029 - 0.0045 \\ 0.221 - 0.227 & 0.9730 - 0.9744 & 0.039 - 0.044 \\ 0.0048 - 0.014 & 0.037 - 0.043 & 0.9990 - 0.9992 \end{pmatrix}. \tag{1.19}$$

To reflect this, the standard parameterisation of $\mathbf{V}_{CKM}$ recommended by the Particle Physics Data group is

$$\mathbf{V}_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{-i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{-i\delta} & s_{23}c_{13} \\ s_{12}c_{23} - c_{12}c_{23}s_{13}e^{-i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{-i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.20}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. θ_{ij} are the three real parameters and δ is the single phase. Notice the phase appears on the smallest elements, reflecting the fact that *CP* violation is small in the SM. Having a non-zero (or non-integral multiple of π) phase does still not guarantee *CP* violation as a couple of further conditions must be met. Degeneracy in any two quark masses of the *same* type allows yet another quark field rotation similar to Eq.(1.18) that can remove the remaining phase. Secondly the phase can be removed if *any* of the three real angles are equal to 0 or $\pi/2$, for then some of the elements above would become zero and the respective quark fields would 'decouple'. A second parameterisation, and the one adopted herein, is that of the Wolfenstein parameterisation [16]

$$\mathbf{V}_{CKM} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{1.21}$$

where A, ρ and η are all real parameters of order unity and $\lambda \sim 0.22$. This parameterisation again reflects the hierarchy of the experimentally obtained magnitudes.

1.4 *CP* violation in the *b*-sector

Having discussed *CP* violation in the SM, the connection to the physics of B mesons can be made. As stated earlier $\mathbf{V}_{CKM}$ is defined to be unitary. This condition implies nine orthonormal relations, six of which can be represented as triangles in the complex plane. The relation of relevance to *CP* violation in B mesons is given by

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.22}$$

Each one of these terms (each describing a side of the triangle) is of order λ^3, therefore we can expect the angles of the triangle to be of order unity and hence the decays of b quarks to contain weak phases of order unity. Due to these large weak phases *CP* violation is expected to large in the b-sector. The relationship between the sides and angles of the unitarity triangle is shown pictorially in Figure 1.1 (a), with the triangle rescaled and oriented in the complex plane in (b). This is achieved by dividing the side lengths by $|V_{cd}V_{cb}^*|$ and choosing a phase convention such that $V_{cd}V_{cb}^*$ is real.

The three angles of this triangle α, β and γ are given by

$$\alpha = arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right],\ \beta = arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right],\ \gamma = arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \tag{1.23}$$

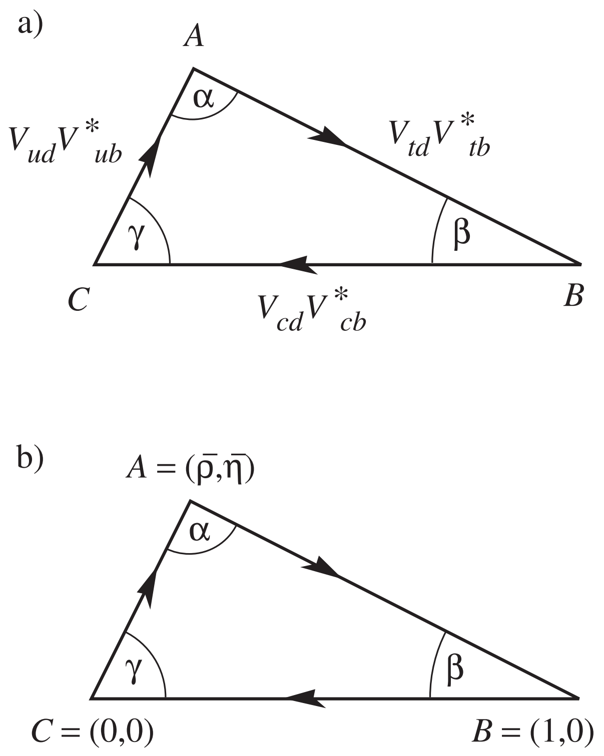


Figure 1.1: The unitarity triangle relevant to the decays of B mesons. Figure (b) shows the triangle rescaled with a phase convention such that $V_{cd}V^*_{cb}$ is real.

These three physical parameters can each be measured independently (albeit indirectly), and thus we can overconstrain the unitarity property of the triangle. If the three angles do not add up to π, the presence of new physics in the decays of B mesons can be inferred. Relating these angles to experimentally observable quantities is not trivial however, and is discussed below. All related measurements, unsurprisingly, involve the analysis of CP-violating processes.

We move now to discuss the three types of processes that can give rise to CP violation in B decays. They are, in no particular order,

i. CP violation in decay or 'direct CP violation'. This occurs for both neutral and charged B mesons and arises when an amplitude for a decay is not equal to its CP-conjugate process.

ii. CP violation in mixing or 'indirect CP violation'. This occurs only for neutral B decays and when the final state is not a CP eigenstate.

iii. *CP* violation in the interference of decays with and without mixing. This occurs when a final state is common to both B and $\overline{B}$ mesons.

1.4.1 Direct *CP* violation

Direct *CP* violation arises when the magnitude of an amplitude for a process $B \to f$ is not equal to its *CP*-conjugate process $\overline{B} \to \bar{f}$. Writing the amplitudes for the two processes as $\mathcal{A} = \langle f|\mathcal{H}|B\rangle$ and $\overline{\mathcal{A}} = \langle \bar{f}|\mathcal{H}|\overline{B}\rangle$ we have

$$\left|\frac{\mathcal{A}}{\overline{\mathcal{A}}}\right| \neq 1 \quad \Rightarrow \quad CP \text{ violation} \tag{1.24}$$

These amplitudes can be expressed as a coherent sum of amplitude contributions a_i, each of which carry two types of phase. One is a *CP*-invariant phase δ_i, commonly called the 'strong phase' due to its usual origin. The other is a *CP*-violating phase ϕ_i known as the 'weak phase' due to the only allowed source of *CP* violation in the SM. The amplitudes are then

$$\mathcal{A} = \sum_i a_i e^{i(\delta_i+\phi_i)}; \quad \overline{\mathcal{A}} = \textstyle\sum_i a_i e^{i(\delta_i-\phi_i)}. \tag{1.25}$$

Equality of the factors a_i in both the B and $\overline{B}$ amplitudes is a consequence of the *CP* and time-reversal invariance of the strong interaction. For direct *CP* violation to occur, there must be at least two differing amplitude components for both $\mathcal{A}$ and $\overline{\mathcal{A}}$ which differ in *both* ϕ and δ. In the case of just two contributing amplitudes, the necessity of two differing phases for *CP* violation is well illustrated in Figure 1.2 (though is of course trivial to derive also). Here we have defined $\delta_0 = \phi_0 = 0$ for simplicity. To observe direct *CP* violation, the lengths of $\mathcal{A}$ and $\overline{\mathcal{A}}$ have to differ. As can be seen this can only be achieved if δ_1 and ϕ_1 are non-zero. To experimentally resolve such differences the differing amplitude components must be of a *comparable* magnitude also - having either a_0 or a_1 much larger than the other will make the triangle too flat to show up appreciable *CP* asymmetries. The physical processes that are thought to give rise to the component amplitudes are discussed below.

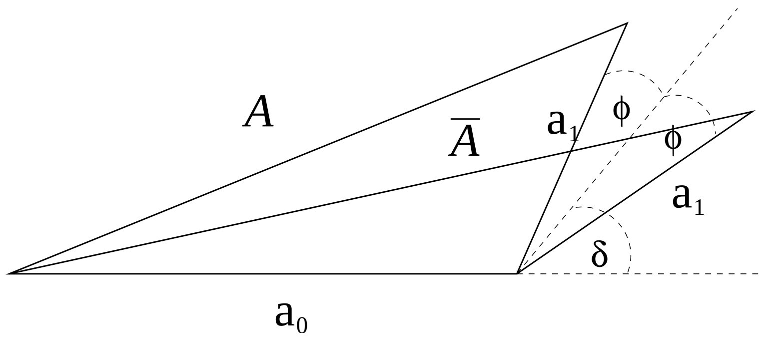


Figure 1.2: Illustration of the need for differing strong and weak phases between the amplitude components to observe direct *CP* violation. To observe direct *CP* violation the lengths $\mathcal{A}$ and $\overline{\mathcal{A}}$ must differ, thus both the *CP*-invariant phase δ and *CP*-violating phase ϕ must be non-zero. Additionally the lengths of the component amplitudes must be of a similar size for experimental observation of *CP* violation.

Direct *CP* violation is the only kind available to charged B mesons and is the easiest to measure experimentally. In neutral B mesons it can occur but may compete with the other forms of *CP* violation and so is less straightforward to untangle. Practically one measures the asymmetry

$$a_f = \frac{\Gamma(B \to f) - \Gamma(\overline{B} \to \bar{f})}{\Gamma(B \to f) + \Gamma(\overline{B} \to \bar{f})} \tag{1.26}$$

which is related to the amplitudes by

$$a_f = \frac{1 - |\overline{\mathcal{A}}/\mathcal{A}|^2}{1 + |\overline{\mathcal{A}}/\mathcal{A}|^2} \tag{1.27}$$

There has been recent observation of direct *CP* violation in the neutral B decay $B^0 \to K^+\pi^-$, at the level of around 11% [17]. This has been observed independently by both *BABAR* [18] and *BELLE* [19].

1.4.2 *CP* violation in mixing - 'indirect' *CP* violation

In discussing neutral B mesons, care must be taken to properly define what we mean. Firstly there are *flavour* eigenstates, which are states with defined valence quark content and are the eigenstates of the strong force. Explicitly $|B^0\rangle = |d\bar{b}\rangle$ and $|\overline{B^0}\rangle = |\bar{d}b\rangle$, with the convention that $CP|B^0\rangle = |\overline{B^0}\rangle$ and $CP|\overline{B^0}\rangle = |B^0\rangle$.

Secondly there are the eigenstates of the free Hamiltonian, which describe states with definite mass and time evolution, and which differ from the flavour eigenstates. B meson flavour eigenstates share common decay channels. Because of this, their time evolution will be such they oscillate from one flavour eigenstate to the other. The dominant oscillation amplitudes are through second order weak interactions of the form illustrated in Figure 1.3.

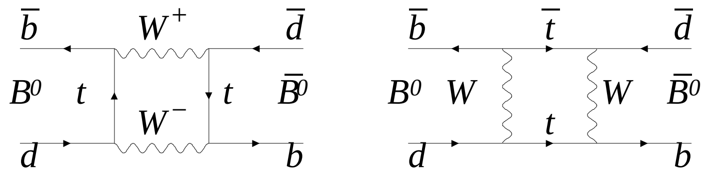


Figure 1.3: The dominant transition diagrams for $B^0 \to \overline{B^0}$ mixing.

We can write an arbitrary neutral B meson state as a linear superposition of flavour eigenstates: $|B(t)\rangle = a|B^0\rangle + b|\overline{B^0}\rangle$. This state vector satisfies the time-dependent Schrödinger equation [13]:

$$i\frac{\partial}{\partial t}|B(t)\rangle = \left[\boldsymbol{M} - i\frac{\boldsymbol{\Gamma}}{2}\right]|B(t)\rangle, \tag{1.28}$$

where

$$\boldsymbol{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{22} \end{pmatrix} \tag{1.29}$$

are Hermitian matrices. The Hamiltonian is thus not Hermitian, and the two eigenstates are not necessarily orthogonal. By CPT invariance, $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$; if CP were conserved

$$\mathrm{Im}(M_{12}) = \mathrm{Im}(\Gamma_{12}) = 0. \tag{1.30}$$

The eigenstates of the Hamiltonian can be written as

$$\begin{aligned} |B_L\rangle &= p|B^0\rangle + q|\overline{B^0}\rangle &= |B_L(0)\rangle e^{-it(M_L - i\frac{\Gamma_L}{2})} \\ |B_H\rangle &= p|B^0\rangle - q|\overline{B^0}\rangle &= |B_H(0)\rangle e^{-it(M_H - i\frac{\Gamma_H}{2})}, \end{aligned} \tag{1.31}$$

with L and H denoting light and heavy eigenstates respectively. p and q are in general complex, but are subject to the normalisation condition $|p|^2 + |q|^2 = 1$. We write the mass and width differences as $\Delta m_B \equiv M_H - M_L$ and $\Delta\Gamma_B \equiv \Gamma_H - \Gamma_L$, and noting that the eigenvalues from the Schrödinger equation are $(M_L - i\frac{\Gamma_L}{2})$ and $(M_H - i\frac{\Gamma_H}{2})$, we get

$$(\Delta m_B)^2 - \frac{1}{4}(\Delta\Gamma_B)^2 = 4(|M_{12}|^2 - \frac{1}{4}|\Gamma_{12}|^2) \tag{1.32}$$

$$(\Delta m_B)(\Delta\Gamma_B) = 4\mathrm{Re}(M_{12}\Gamma_{12}^*) \tag{1.33}$$

such that

$$\frac{q}{p} = -\sqrt{\frac{(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)}{(M_{12} - \frac{i}{2}\Gamma_{12})}} = -\frac{\Delta m_B - \frac{1}{2}\Delta\Gamma_B}{2(M_{12} - \frac{1}{2}\Gamma_{12})} = -\frac{2(M_{12}^* - \frac{1}{2}\Gamma_{12}^*)}{\Delta m_B - \frac{1}{2}\Delta\Gamma_B} \tag{1.34}$$

For CP violation in mixing, the condition in Eq.(1.30) implies

$$\left|\frac{q}{p}\right| \neq 1 \tag{1.35}$$

Practically one can measure the time-dependent asymmetry

$$a_f = \frac{\Gamma(\overline{B}^0(t) \to l^+\nu X) - \Gamma(B^0(t) \to l^-\bar{\nu}X)}{\Gamma(\overline{B}^0(t) \to l^+\nu X) + \Gamma(B^0(t) \to l^-\bar{\nu}X)} \tag{1.36}$$

which in terms of $|q/p|$ is given by

$$a_f = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \tag{1.37}$$

However, this asymmetry is expected to be small (of the order of 10^{-2} [20]) and the theoretical uncertainties in calculating Γ_{12} and M_{12} mean relating this quantity to CKM parameters is difficult, even if measured with precision.

1.4.3 *CP* violation in the interference of decays with and without mixing

Consider a final state f_{CP} which is an eigentstate of CP and is available to both B^0 and $\overline{B}^0$. Expressing the amplitudes for the processes as $\mathcal{A}(B^0 \to f_{CP})$ and

$\overline{\mathcal{A}}(\overline{B^0} \to f_{CP})$ we write the *CP* violating quantity of interest as

$$\lambda = \eta_f \frac{q}{p} \frac{\overline{\mathcal{A}}}{\mathcal{A}} \tag{1.38}$$

where η_f is the *CP* eigenvalue (± 1) of the final state f. For *CP* conservation $|q/p| = |\overline{\mathcal{A}}/\mathcal{A}| = 1$ with the relative phase between the q/p and $\overline{\mathcal{A}}/\mathcal{A}$ vanishing. For *CP* violation we have

$$\lambda \neq \pm 1. \tag{1.39}$$

This condition can be brought about by both *CP* violation in mixing and in decay. However to a good approximation in most cases it is still true that $|q/p| = 1$ and $|\overline{\mathcal{A}}/\mathcal{A}| = 1$, and *CP* violation arises from a non-vanishing phase difference between the two ratios. In this case, λ acquires a phase such that

$$|\lambda| = 1, \quad \mathrm{Im}\lambda \neq 0. \tag{1.40}$$

This type of *CP* violation is called '*CP* violation in the interference of decays with and without mixing'. In practice one measures the time-dependent quantity

$$a_{f_{CP}} = \frac{\Gamma(\overline{B}^0(t) \to f_{CP}) - \Gamma(B^0(t) \to f_{CP})}{\Gamma(\overline{B}^0(t) \to f_{CP}) + \Gamma(B^0(t) \to f_{CP})} \tag{1.41}$$

which can be re-expressed as

$$a_{f_{CP}} = \frac{(1 - |\lambda|^2)\cos(\Delta m_B t) - 2Im\lambda\sin(\Delta m_B t)}{1 + |\lambda|^2}. \tag{1.42}$$

In the case where there is only one weak amplitude contributing to this process, λ becomes a pure phase and the asymmetry becomes considerably more simple

$$a_f(t) \simeq \eta_f \sin 2\beta \sin \Delta m t. \tag{1.43}$$

This is the case for certain decays of the form $b \to c\overline{c}s$, as discussed in section 1.4.4.

We move now to discuss how the CKM angles are extracted experimentally.

1.4.4 Measuring β

The angle β is the easiest of the three CKM angles to measure experimentally. This is because the quantity $\sin 2\beta$ can be readily extracted from the decays of neutral B mesons to final states involving charmonium and a neutral kaon. Charmonium decays present a clean experimental signature, are relatively abundant in B decays and the theoretical calculations involved in describing the amplitudes have small model dependent uncertainties. The two principal Feynman diagrams for the processes of these 'Type I' decays are shown in Figure1.4. The small theoretical uncertainty chiefly arises from the tree and penguin diagrams having the same weak phase to first order in their matrix elements. This feature reduces the model dependence of the calculation as no additional assumptions about factorisation, colour suppression or final state phases are built in. For this reason these channels, and in particular the decay to $J/\psi K_S$ are often called "Golden Channels". The time-dependent

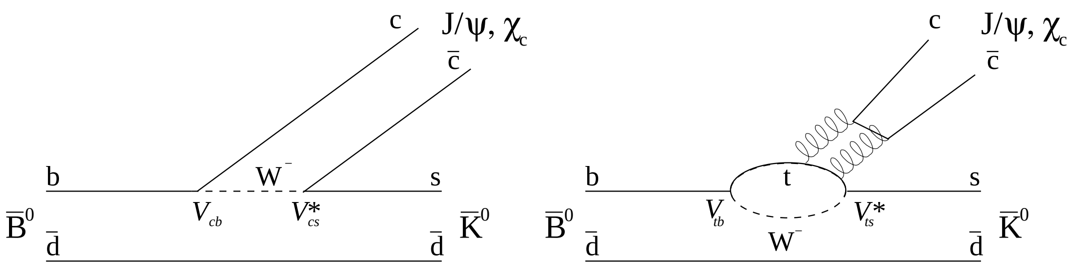


Figure 1.4: Tree and penguin diagrams contributing to the "Golden Channel" $\sin 2\beta$ process.

asymmetry for this type of decay is given by Eq.(1.43).

There are other methods to extract $\sin 2\beta$, such as Dalitz plot analyses[2] of $B \to D$ decays. The current world-average measurement for $\sin 2\beta$ is 0.739 ± 0.048 [17].

1.4.5 Measuring α

Processes of the form $b \to du\bar{u}$ can in principle be used to measure α. In particular the decay $B^0 \to \pi^+\pi^-$ lends itself to the measurement of $\sin 2\alpha$ in an analogous

[2] These are discussed in sections 1.7 and 1.8.

way to the Golden Channel process for $\sin 2\beta$. Unlike the $\sin 2\beta$ case however, the

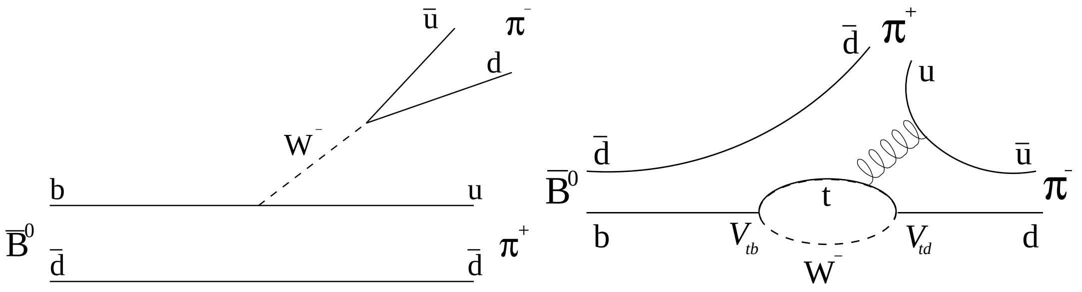


Figure 1.5: Tree and penguin diagrams contributing to decay $B^0 \rightarrow \pi^+\pi^-$ used in the measurement of $\sin 2\alpha$.

weak phases present in penguin and tree amplitudes are no longer the same and penguin contributions have been found to be non-negligible [21], thereby spoiling a clean extraction of the angle [22]. The fact that the $\pi^+\pi^-$ process is contaminated by penguins leads us to consider complementary modes that may also yield information on α. Though no channel offers a clean extraction, most suffering the same fate of non-negligible penguin contributions, that is not to say α is beyond reach. The technique known as isospin analysis offers a means by which to distinguish the tree processes from the penguin ones. This tool exploits a fact that for certain channels there is at least one amplitude describing an isospin transition that can be reached *only* by a tree process. By isolating such processes one may untangle the tree-only contribution and thereby remove the penguin pollution. This technique is only applicable to certain select modes, such as $\pi\pi$, $K\pi$, $\rho\pi$, $\rho\rho$, *etc.* [23]. The amplitudes for the differing ΔI transitions of the component amplitudes are labelled $A_{\Delta I, I_f}$. In the case of the decay $B \rightarrow \rho\pi$, which is of interest here, the isospin-decomposed amplitudes which contribute are summarised in table 1.1. The penguins contributing to these modes are necessarily $\Delta I = \frac{1}{2}$ transistions[3].

Writing these amplitudes explicitly in terms of their tree (T) and penguin (P) contributions along with the weak phase α we get

[3]The penguins describe the processes $b \rightarrow q\bar{q}d$ with the quark-antiquark pair produced by gluons which carry $I = 0$. The only change in isospin number therefore comes from the d quark.

Decay Amplitudes									
$\mathcal{A}(B^+ \to \rho^+\pi^0)$	$=$	$\frac{1}{2}\sqrt{\frac{3}{2}}A_{3/2,2}$	$-$	$\frac{1}{2}\sqrt{\frac{1}{2}}A_{3/2,1}$	$+$	$\sqrt{\frac{1}{2}}A_{1/2,1}$		
$\mathcal{A}(B^+ \to \rho^0\pi^+)$	$=$	$\frac{1}{2}\sqrt{\frac{3}{2}}A_{3/2,2}$	$+$	$\frac{1}{2}\sqrt{\frac{1}{2}}A_{3/2,1}$	$-$	$\sqrt{\frac{1}{2}}A_{1/2,1}$		
$\mathcal{A}(B^0 \to \rho^+\pi^-)$	$=$	$\frac{1}{2}\sqrt{\frac{1}{3}}A_{3/2,2}$	$-$	$\frac{1}{2}A_{3/2,1}$	$+$	$\frac{1}{2}A_{1/2,1}$	$-$	$\sqrt{\frac{1}{6}}A_{1/2,0}$
$\mathcal{A}(B^0 \to \rho^-\pi^+)$	$=$	$\frac{1}{2}\sqrt{\frac{1}{3}}A_{3/2,2}$	$+$	$\frac{1}{2}A_{3/2,1}$	$-$	$\frac{1}{2}A_{1/2,1}$	$-$	$\sqrt{\frac{1}{6}}A_{1/2,0}$
$\mathcal{A}(B^0 \to \rho^0\pi^0)$	$=$	$\sqrt{\frac{1}{3}}A_{3/2,2}$	$+$	$\sqrt{\frac{1}{6}}A_{1/2,0}$				

Table 1.1: Isospin decomposition of $B \to \rho\pi$ in terms of the isospin amplitudes $A_{\Delta I, I_f}$.

$$\begin{array}{rclcl}
\mathcal{A}(B^0 \to \rho^+\pi^-) & \equiv & A^{+-} & = & e^{-i\alpha}T^{+-} + P^{+-}, \\
\mathcal{A}(B^0 \to \rho^-\pi^+) & \equiv & A^{-+} & = & e^{-i\alpha}T^{-+} + P^{-+}, \\
\mathcal{A}(B^0 \to \rho^0\pi^0) & \equiv & A^{00} & = & e^{-i\alpha}T^{00} + P^{00}, \\
\frac{q}{p}\mathcal{A}(\overline{B}^0 \to \rho^+\pi^-) & \equiv & \frac{q}{p}\overline{A}^{+-} & = & e^{+i\alpha}T^{-+} + P^{-+}, \\
\frac{q}{p}\mathcal{A}(\overline{B}^0 \to \rho^-\pi^+) & \equiv & \frac{q}{p}\overline{A}^{-+} & = & e^{+i\alpha}T^{+-} + P^{+-}, \\
\frac{q}{p}\mathcal{A}(\overline{B}^0 \to \rho^0\pi^0) & \equiv & \frac{q}{p}\overline{A}^{00} & = & e^{+i\alpha}T^{00} + P^{00},
\end{array}$$

and for the charged modes

$$\begin{array}{rcl}
\sqrt{2}\mathcal{A}(B^+ \to \rho^+\pi^0) & = & e^{-i\alpha}T^{+0} + P^{+-} - P^{-+}, \\
\sqrt{2}\mathcal{A}(B^+ \to \rho^0\pi^+) & = & e^{-i\alpha}(T^{+-} + T^{-+} + 2T^{00} - T^{+0}) - P^{+-} + P^{-+},
\end{array} \tag{1.44}$$

with

$$P^{00} = \sqrt{\frac{1}{6}}A_{1/2,0} = -\frac{1}{2}(P^{+-} + P^{-+}). \tag{1.45}$$

Although in principle the three tree and three penguin contributions to the neutral B amplitudes can be untangled from the neutral relations alone [24], the inclusion of the two charged B modes adds only one new tree component T^{+0}. Thus measurement of the charged B decays gives independent information about the penguin pollution.

1.4.6 Measuring γ

The angle γ is regarded as the most difficult to extract experimentally. In addition, observable quantities which depend on γ also depend on other ill-known parameters, making theoretical untanglement of experimental results highly non-trivial. There are several strategies devised to determine γ experimentally, one of which is based on theoretical amplitude relations between decays of B mesons to final states composed of a K and a π [25] [26]. These 'triangle' relations are derived principally from $SU(3)$ flavour symmetry although they have certain dynamical assumptions built in, such as absence of annihilation topologies, electroweak penguins, etc. One important feature of these relations is that only the branching fractions have to be measured. However it is now believed that assumptions such as the absence of electroweak penguins are no longer valid [27]. Additionally, final state interactions are believed to be non-negligible for $K\pi$-like modes, thereby changing the strong phases by an unknown amount. All these features compound to make the triangular relations naive at best.

1.4.7 State of the art

The present knowledge about the CKM angles is summarised in Figure 1.6. This is formed from all current, world-averaged measurements on $\sin 2\beta$. In addition there are other measurements which act to constrain the apex of the triangle such as Δm_B from B_d-mixing and ε_K from K-mixing which are included in this plot. The coloured regions show the confidence levels obtained from all the various measurements [28]. Note there are as yet no inputs from measurements related to α and γ in this plot.

1.5 Charmless B physics

The decays of B mesons to final states with the absence of charm quarks present a rich and complex phenomenology. This is largely due to the presence of the second order penguin diagrams, which can have magnitudes comparable to some of the tree processes. The Feynman diagrams for the amplitudes available to charmless decays are shown in Figure 1.7. The labelling is as follows: T denotes the colour-favoured tree, C the colour-suppressed tree (suppressed by a factor of around three compared

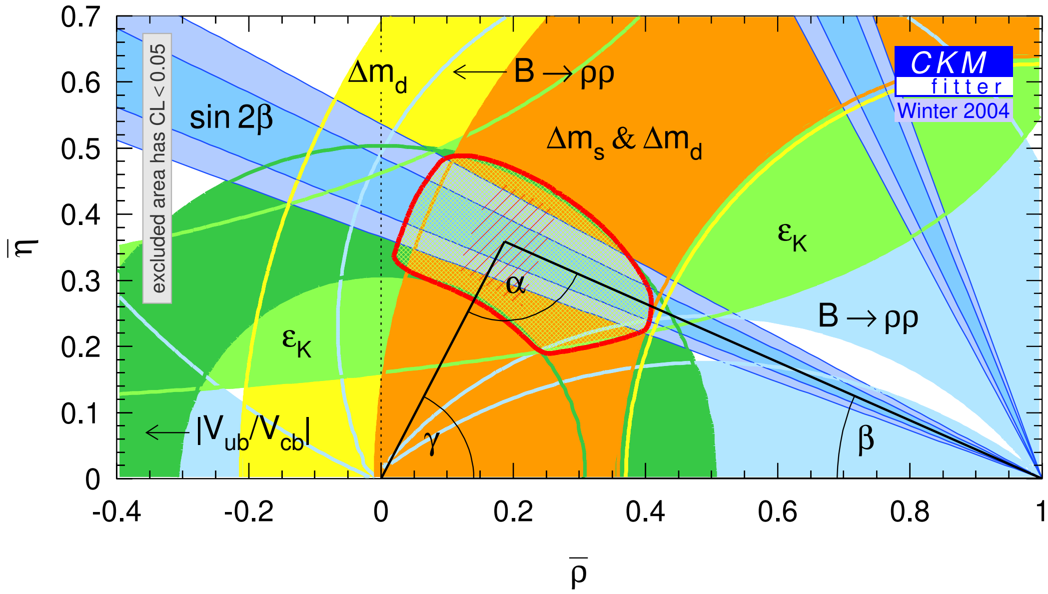


Figure 1.6: World average confidence levels of the unitarity triangle in the $\overline{\rho},\overline{\eta}$ plane. The coloured regions show the confidence levels obtained from different measurements. This plot was produced with [28].

to T), P is the QCD penguin, A the annihilation, P_{EW} the colour-favoured electroweak penguin, P_{EW}^{C} the colour-suppressed electroweak penguin, EP the penguin exchange and EP_{EW}^{C} the colour-suppressed electroweak penguin exchange. Each penguin can have one of three quarks (u, c, t) in the loop, though any one of these separate amplitudes can always be factored out using the triangle relations derived from the unitarity of the CKM matrix, yielding only two independent diagrams per penguin. The hierarchy assumption in most models of charmless B physics is that the QCD penguin amplitudes are of the same order of magnitude as the tree amplitudes ($P \sim T$) within CKM factors[4], electroweak penguin amplitudes are roughly one order of magnitude smaller than QCD penguins, while annihilation amplitudes are very small and in most cases are ignored. The presence of sizeable penguin amplitudes allow for the possibility of seeing large direct CP asymmetries, by providing the required differing strong and weak phases from the tree amplitude. However if the penguin amplitudes are too large compared to the tree, observing direct CP becomes harder as discussed earlier.

[4]For example, in the diagrams of Figure 1.7 where the final state is a kaon, $|TV_{ub}^{*}V_{us}|/|PV_{tb}^{*}V_{ts}| \sim 0.2$ [27]. In the case of $B \rightarrow \pi^{+}\pi^{-}$ the tree is dominant, though QCD penguins are still significant.

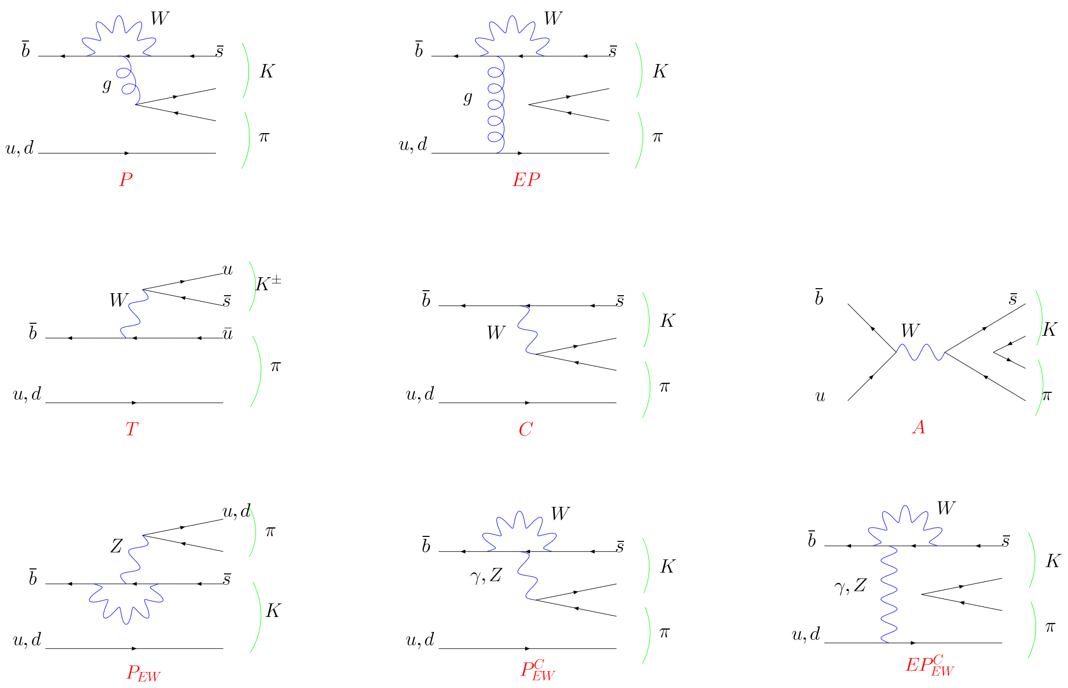


Figure 1.7: Tree and penguin diagrams contributing to charmless B decays. Figure is taken from [27].

1.5.1 Approximate theories

In describing the physics of B decays, one must deal with both the bound states of the B mesons and with hadronic final state interactions which, though in principle governed by QCD, are very often in the non-perturbative regime and are therefore notoriously difficult, if not impossible, to calculate. To overcome these difficulties, techniques have been developed whereby predictions for physically significant quantities such as branching fractions and CP asymmetries can be made under certain assumptions. The details of these techniques are not given here, rather only a few words outlining the underlying principles.

Effective theories

Effective theories are based on the principle that any given process has a finite number of degrees of freedom that are of physical significance. By appealing to such things as the kinematics of the decay, one may reasonably remove from the theory the intermediate states of high virtuality by reabsorbing them into the degrees of freedom of the process, thus introducing 'effective' interactions which can be approximated. Examples of effective theories are the operator product expansion (OPE), which expands propagators in terms of the virtuality of the exchanged particles, heavy-quark effective theory (HQET) and chiral perturbation theory (ChPT).

Factorisation

Factorisation is a method that approximates incalculable QCD matrix elements of the form

$$\langle h_1 h_2 | \mathcal{O} | B \rangle \tag{1.46}$$

with the product of matrix elements

$$\langle h_1 | \mathcal{J}_1 | B \rangle \langle h_2 | \mathcal{J}_2 | 0 \rangle \tag{1.47}$$

where the currents $\mathcal{J}_{1,2}$ are formed from empirically known quantities such as form factors and branching ratios. The assumptions are that non-perturbative effects are negligible in part due to the heavy mass of the b quark, and the light mass of the final hadron h_2 which typically is produced with large momentum.

Light flavour symmetry

Light flavour or flavour-$SU(3)$ symmetry is frequently used to mitigate the effects of hadronic uncertainties in charmless B decays. It is based on the symmetries of the physics in the limit of vanishing u, d and s quark masses, whereby the three states become degenerate with respect to the strong interaction. For example the amplitudes for charged B decays to $\pi\pi$, $K\pi$ can be related [25] [26] by the expression

$$A(B^- \to \pi^- \overline{K}^0) + \sqrt{2}A(B^- \to \pi^0 K^-) = \sqrt{2}\left(\frac{V_{us}}{V_{ud}}\right) A(B^- \to \pi^- \pi^0)\{1 + \delta_{SU(3)}\} \tag{1.48}$$

where $\delta_{SU(3)}$ is the degree of $SU(3)$ violation. $\delta_{SU(3)}$ is usually related phenomenologically to the ratio of kaon and pion form factors f_K/f_π, though such estimates are reliant on factorisation assumptions.

1.5.2 Long distance rescattering effects for $K\pi$ channels

There has been discussion in the literature [29] that amplitude contributions to allow direct CP violation in $K\pi$ modes may arise from a source other than the $\Delta I = 0$ penguin amplitudes. This source is that of a long-distance final state interaction Hamiltonian. This interaction is not understood well enough to be reliably calculated, though in principle could give rise to large direct CP asymmetries that are unique and independent to each final state. Indeed such an interaction would spoil the $SU(3)$ amplitude relations used to extract γ which implicitly assumed this effect was small or absent. However, theorists are in disagreement as to the extent of this effect thus it is important to measure direct CP violating asymmetries in as many $K\pi$ modes as possible.

1.6 Theoretical predictions for the decays $B^{\pm} \to K^{*\pm}\pi^0$ and $B^{\pm} \to \rho^{\pm}\pi^0$

There have been two recent theoretical predictions for the, as yet unmeasured, branching fraction $\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^0)$ [30] [31]. The first by Beneke and Neubert, derived from QCD factorization, gives $\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^0) = (3.3\,^{+1.1+1.0+0.6+4.4}_{-1.0-0.9-0.6-1.4}) \times 10^{-6}$, where the four errors correspond to (1) uncertainties from the variation of CKM parameters, (2) variation of renormalization scale, quark masses, decay constants, and form factors, (3) uncertainty from expansion of light-cone distribution amplitudes, and (4) the estimate of power corrections. The value of γ assumed is $70°$. The second prediction of Chiang and Gronau relies on assumptions of isospin and SU(3) flavour symmetry and expects branching fractions between $(15.0^{+3.3}_{-2.8}) \times 10^{-6}$ and $(22.1^{+4.2}_{-5.1}) \times 10^{-6}$, depending on the value of the CKM angle γ. The only experimental result for this branching fraction published to date is an upper limit of $\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^{0}) < 31 \times 10^{-6}$, where the fit had a central value of $(7.1^{+11.4}_{-7.1} \pm 1.0) \times 10^{-6}$ and a yield of 2.6 events, using the $K^{\pm}\pi^{0}\pi^{0}$ final state [32]. A measurement of this mode would hope to resolve the discrepancy of the predictions.

These papers also make predictions for the *CP*-averaged branching fractions of the already measured [17] decay $B^{\pm} \to \rho^{\pm}\pi^{0}$. They are respectively $\mathcal{B}(B^{\pm} \to \rho^{\pm}\pi^{0}) = (14.0\,^{+6.5+5.1+1.0+0.8}_{-5.5-4.3-0.6-0.7}) \times 10^{-6}$ (for $\gamma = 70°$) and in the range $([10.1 \to 11.8]^{+1.6}_{-1.8}) \times 10^{-6}$ depending on the value of γ assumed. These are to to be compared with the experimentally measured value of $(12.0 \pm 2.0) \times 10^{-6}$, which is the average of the results obtained by *BELLE* and *BABAR*.

Additionally the papers make predictions for the direct *CP* asymmetries for the two decay modes. The paper of Beneke and Neubert predicts $\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^{0}) = 0.087^{+0.021+0.050+0.029+0.417}_{-0.026-0.043-0.034-0.442}$. and $\mathcal{A}_{CP}(B^{\pm} \to \rho^{\pm}\pi^{0}) = -0.040^{+0.012+0.018+0.004+0.175}_{-0.012-0.022-0.004-0.177}$. The paper of Chiang and Gronau predicts the *CP* asymmetries in the ranges $\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^{0}) = -0.04^{+0.03}_{-0.02} \to 0.01 \pm 0.05$ and $\mathcal{A}_{CP}(B^{\pm} \to \rho^{\pm}\pi^{0}) = -0.01 \pm 0.06 \to 0.16^{+0.04}_{-0.05}$, depending on the value of γ assumed. The experimental measurement for the direct *CP* asymmetry in the decay $B^{\pm} \to \rho^{\pm}\pi^{0}$ is 0.16 ± 0.13 [17], again averaged over the *BABAR* and *BELLE* results.

A refinement of the $B^{\pm} \to \rho^{\pm}\pi^{0}$ branching fraction and *CP* asymmetry measurement would help to constrain the parameter space available to these theoretical models.

1.7 Kinematics of three-body decays

A decay process $\langle \phi_1, \phi_2, ...\phi_N|\mathcal{H}|\phi_i\rangle$ for a spinless particle can be described fully in terms of the N four-momenta of the final state. This provides 4N degrees of freedom

(DOF). However, these four-vectors are each constrained by the N masses of the final state particles, so N degrees of freedom in the interaction are immediately lost, leaving 3N. In addition, the components of each final state four-momentum must sum to those of the parent by energy-momentum conservation so the process loses four more DOF. Finally, the physics of the decay cannot depend on the orientation in phase space thus the freedom to rotate the whole system in the three Euler angles is not physically significant, leaving a final total of 3N-4-3 degrees of freedom. For N = 2 only two Euler angles can be defined and there are no remaining DOF. In the special case of three-body decays, there are thus only two independent variables and any particular decay can be conveniently described by a single point on a plane.

Recall the relation for the decay width $d\Gamma$ for a three-body decay $D \to abc$ with transition amplitude $\mathcal{A}$ is given by

$$d\Gamma = \frac{|\mathcal{A}|^2}{256\pi^2 M^3} dm_{ab}^2 dm_{ac}^2, \tag{1.49}$$

where M is the mass of the parent and $m_{ab}^2 = (p_a + p_b)^2$, $m_{ac}^2 = (p_a + p_c)^2$ are two Mandelstam variables. Motivated by this relation, a popular (though by no means unique) choice for our two variables are these two Mandelstam variables m_{ab}^2 and m_{ac}^2, thereby defining what is commonly called the *Dalitz plot* after its first documented use [33]. In the absence of an amplitude which depends on m_{ab}^2 and m_{ac}^2, the allowed region in the Dalitz plot is uniformly populated. The Dalitz plot is illustrated graphically in Figure 1.8.

1.8 Resonance kinematics, amplitudes and interference

A resonance in quantum mechanics is essentially a description of a state whose lifetime is large compared to the time needed to create it, or alternatively a state which decays slowly relative to the frequency of oscillation of the wave function $E/\hbar$. There are many different types of resonances in nuclear and particle physics

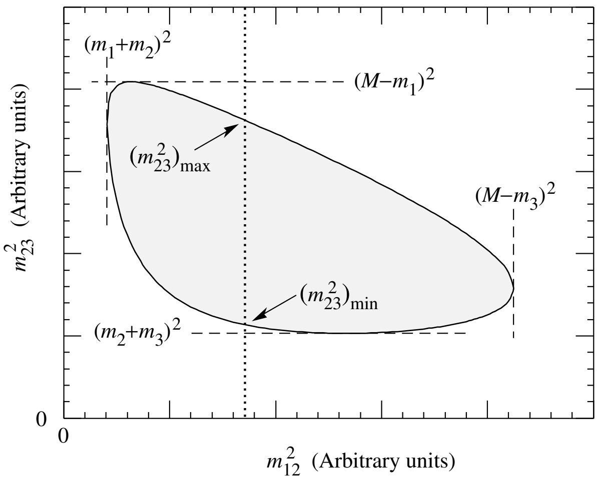


Figure 1.8: Pictorial representation of a Dalitz plot. The shaded area represents the kinematically allowed region. Figure is taken from [15].

and it is indeed fortunate that almost all can be described with the same underlying prescription given by the condition of unitarity[5].

The most prolific example of a resonant amplitude is that of the non-relativistic Breit-Wigner,

$$\mathcal{A}(E) = \frac{-ia\Gamma}{(E - E_R) - i\Gamma/2}, \tag{1.50}$$

where $|\mathcal{A}(E)|^2\delta E$ is the probability of finding the state with an energy in the range $[E,\, E + \delta E]$. Here Γ is the characteristic width and is related to the lifetime of the decaying state by $\Gamma = \hbar/\tau$ and a is a constant that depends on quantum numbers of the resonance such as its spin. Under relativistic conditions, that is conditions *internal* to the resonance that must be described by relativistic quantum mechanics, this probability density typically takes the form

$$|\mathcal{A}(s)|^2 = \frac{m_0\Gamma}{(s - m_0^2)^2 + (m_0\Gamma)^2}, \tag{1.51}$$

where $\sqrt{s}$ is the center of mass energy, m_0 is the the resonance pole, and $\Gamma(m)$ is now a function of the resonance invariant mass $m = \sqrt{s}$. Relativistic descriptions of

[5]See for example [12]

resonances are in general quite diverse. For the $\rho \to \pi\pi$ a common parameterisation for the width is given by

$$\Gamma(m) = \frac{m_0 \Gamma_0}{m} \left(\frac{p_{CM}}{p_{CM0}} \right)^3 \frac{1 + (R p_{CM0})^2}{1 + (R p_{CM})^2}, \tag{1.52}$$

p_{CM} is the momentum of the pion in the ρ centre of mass frame, and the subscript 0 denotes the same quantities at the resonance pole. R is the semi-classical radius of the strongly interacting region. The last term $1 + (Rp_{CM0})^2/1 + (Rp_{CM})^2$ is known as the Blatt-Weisskopf barrier penetration factor.

If a final state is reachable through two or more amplitudes, quantum mechanical interference may occur. In the case of three-body final states, the underlying structure of interfering amplitudes can be extracted in principle from an examination of the Dalitz plane. For instance, consider only two contributing amplitudes $A_1(m_{ab}, \theta_H)$ and $A_2(m_{ab}, \theta_H)$ as functions of the invariant mass of the pair a and b, and helicity angle $\theta_H = \theta_{bc}$. We may separate the amplitudes into components $A(m_{ab}, \theta_H) = M(m_{ab})\Phi(\theta_H)$. If A describes an amplitude proceeding through a resonant state, $M(m_{ab})$ is the resonant lineshape form (eg. Breit-Wigner, Flatte etc.), while $\Phi(\theta_H)$ takes the form of the Legendre polynomial appropriate to the angular momentum of the resonance. For S, P and D waves, the explicit forms of $\Phi(\theta_H)$ are

$$\Phi(\theta_H)_S = 1; \tag{1.53}$$

$$\Phi(\theta_H)_P = \cos\theta_H; \tag{1.54}$$

$$\Phi(\theta_H)_D = \frac{1}{2}(3\cos^2\theta_H - 1); \tag{1.55}$$

The orthogonality of the Legendre polynomials ensures the interference term integrates to zero for resonances with differing angular momenta provided the limits of integration are symmetric. What the interference term does in these cases is shift the distributions of events but not enhance or reduce the *total* number of events. However, if there is an asymmetric detection efficiency over the helicity angle, one must be considerate of the possibility of interference effects when integrating over the helicity.

1.8.1 Quasi two-body analysis

In cases of three-body analyses which expect low signal statistics and large backgrounds and which may suffer from experimental complications (such as poor resolution and ill-defined efficiency distributions over the Dalitz plane), a full Dalitz-plot analysis to extract phase and amplitude information of differing decay modes can be at or beyond the limit of experimental feasibility.

In such instances one may still attempt to make measurements of physically interesting quantites for various resonances in the phase space by adopting what is known as a 'quasi two-body' approach. This essentially involves selecting events within a narrow band around the invariant mass of the resonance, giving consideration to the possibility of interfering amplitudes in the systematic uncertainties on any measurements. This is the strategy adopted in this thesis for both the $B^{\pm} \to \rho^{\pm}\pi^{0}$ and $B^{\pm} \to K^{*\pm}\pi^{0}$ analyses.

Chapter 2

The PEP-II Storage Rings and the BABAR Detector

2.1 Introduction

The PEP-II B factory and *BABAR* experiment constitute a sensitive laboratory for the precise measurement of the CKM unitarity matrix. The facility is designed principally to scrutinize the decays of neutral B mesons to CP eigenstates and make measurements of any manifest time-dependent CP asymmetries. Three essential ingredients of the B factory concept for this purpose are

- High statistics. CP violation may be large in the B sector, but interesting CP violating decays have branching fractions of $10^{-4} - 10^{-6}$ or less. To see statistically significant CP violating effects, hundreds of millions of B decays are required. This is achieved by the exceptional luminosity delivered by PEP-II.
- High signal to background ratio. This is provided by the clean e^+e^- interaction environment.
- Observable B time-evolution. CP violation can be revealed through a difference in decay rates $\Gamma(B \to f) - \Gamma(\overline{B} \to \bar{f})$, where the bar denotes the CP

conjugate state. Reconstructing the final state alone is not enough, as when f is an eigenstate of *CP* one cannot tell from which B it came. One must have some knowledge of the probable flavour of B meson that produced the final state observed. One solution to this problem, which is adopted by both *BABAR*and *BELLE*, is to produce pairs of $b\bar{b}$ mesons in a coherent state, and tag the flavour of the 'other' B by observing its decay. This effectively acts as starting a stopwatch for one knows the flavour of the B at the moment the tagged B decays[1]. However, the B^0 (and $B^{\pm}$) lifetime is $\sim 1.5 \times 10^{-12}$s [15], a period too short to measure directly and thereby differentiate in time one B meson decay from its partner's. The solution adopted by the *BABAR* and *BELLE* experiments is not to differentiate between the decays in time, but rather in space by giving a relativistic boost to the $B\overline{B}$ centre of mass frame (CM) with the use of asymmetric beams, and interpreting the difference in lifetimes from the distance between two decay vertices.

What follows is a brief description of the PEP-II rings and *BABAR* detector.

2.2 The PEP-II Storage Rings

PEP-II consists of two storage rings, the High Energy Ring (HER), sustaining a beam of 9.0 GeV electrons, and the Low Energy Ring (LER), with a beam of 3.1 GeV positrons. The two rings are injected with electrons and positrons produced with the Stanford Linear Accelerator Center (SLAC) and extracted using a dedicated bypass line. When brought to collision these beams produce a CM energy of 10.58 GeV which corresponds to the location of the maximum of the $\Upsilon(4S)$ resonance, the lowest lying $b\bar{b}$ bound state that can freely decay to two B mesons. In addition, B mesons produced in this way are created coherently, an important part of the strategy for measuring *CP* violation at *BABAR*. The $\Upsilon(4S)$ resonance is excited in e^+e^- collisions at this CM energy in approximately 10% of all annihilation collisions [20], and it

[1]Of course the tagged B need not decay first, but in this case the principle still holds only with the stopwatch time running backwards.

Parameters	Design	Typical	Best [34]
Energy HER/LER (GeV)	9.0/3.1	9.0/3.1	9.0/3.1
Current HER (A)	0.75	1.33	1.55
Current LER (A)	2.15	2.08	2.45
# bunches	1658	1500	1588
Luminosity (10^{33} cm^{-2}s^{-1})	3	7.3	9.2
Luminosity (pb/day)	135	627	710.5

Table 2.1: Performance summary of the PEP-II machine.

subsequently decays to $B\overline{B}$ with a frequency greater than 96% [15]. The intentional asymmetry of the beam energies produces a CM frame boost of $\beta\gamma = 0.56$, allowing the B mesons to traverse an average distance of 0.25 mm, far enough to discern time-dependent *CP* asymmetries on a statistical basis. To characterise processes from light quark and leptonic continuum production, PEP-II is run at a CM energy 40 MeV below the $B\overline{B}$ threshold for 12% of the time.

2.2.1 Beam parameters

The HER and LER beam energies are calculated from the applied magnetic bending field strength and the average deviations of the accelerating frequencies from their central values. These values are sampled every 5 seconds and corrected for a systematic bias in order to give the *BABAR* measured CM energy. Table 2.1 summarises some of the PEP-II design parameters and their typical values during present running.

As can be seen, PEP-II has surpassed its design goals both in terms of instantaneous and daily integrated luminosity. In addition, the PEP-II machine control group have been setting new records nearly every month during operational periods.

2.2.2 Interaction region

The two beams are brought to collision at a single point on the PEP-II rings designated Interaction Region 2 (IR2), where the *BABAR* detector sits. The LER beam is brought down into the plane of the HER and aligned to interact using a pair of dipole magnets (B1) each offset by 21cm from the IP, well within the *BABAR* detector. Also within the *BABAR* detector are two quadrupole magnets (Q1) used for final focus of the LER beam. Both B1 and Q1 are subject to *BABAR*'s 1.5 T solenoidal magnetic field, and are made from samarium-cobalt. The proximity to the detector of these magnets is necessary to avoid secondary collisions of the bunches away from the IP, which would otherwise occur about 60 cm further downstream; however the presence of the magnets is the limiting factor to the *BABAR* acceptance angle. Q4 and Q5, used for HER beam focusing and situated outside the *BABAR* detector, are iron septum electro-magnets. The vacuum chamber of the IR is a 25 mm radius water-cooled beryllium pipe, evacuated to a pressure of $\simeq$ 1 ntorr. The magnets around the IP are shown schematically in Figure 2.1. The IR, beam pipe, magnets and some detector elements are enclosed by a 4.5m long support pipe.

2.3 The *BABAR* detector

The *BABAR* detector [35] (Figure 2.2) is engineered to study the interesting physics presented by PEP-II. It is composed of 5 nested subsystems, in order from innermost outwards: the Silicon Vertex Tracker (SVT), the Drift Chamber (DCH), the Detector of Internally Reflected Cherenkov light (DIRC), the Electromagnetic Calorimeter (EMC) and the Instrumented Flux Return (IFR). The first four subsytems are immersed in a coaxial 1.5 T superconducting solenoidal magnetic field, which is used to identify the charge of tracks while the IFR provides the field return path. These subsystems, together with the trigger system are described further in the following sections.

The right handed Cartesian coordinate system of *BABAR* is defined such that z is in the direction roughly colinear with the HER (the alignment is actually with respect to the DCH), x points radially outwards from the center of the PEP-II rings and y points vertically upwards.

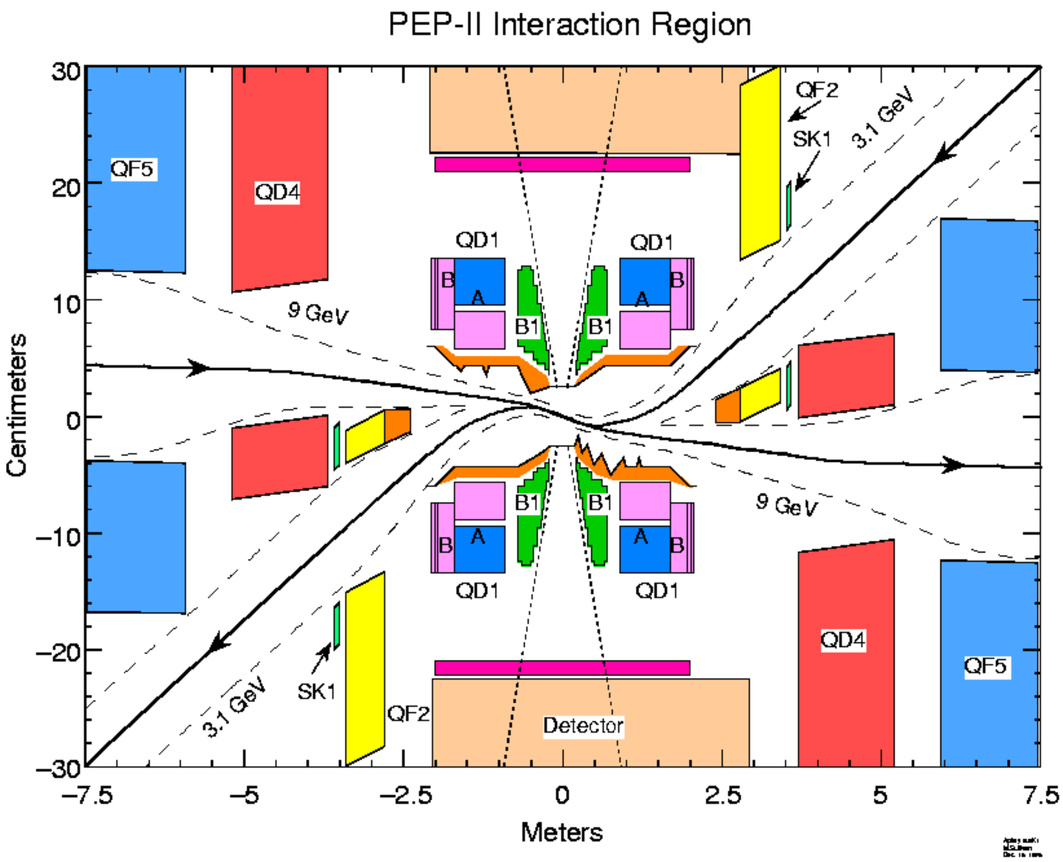


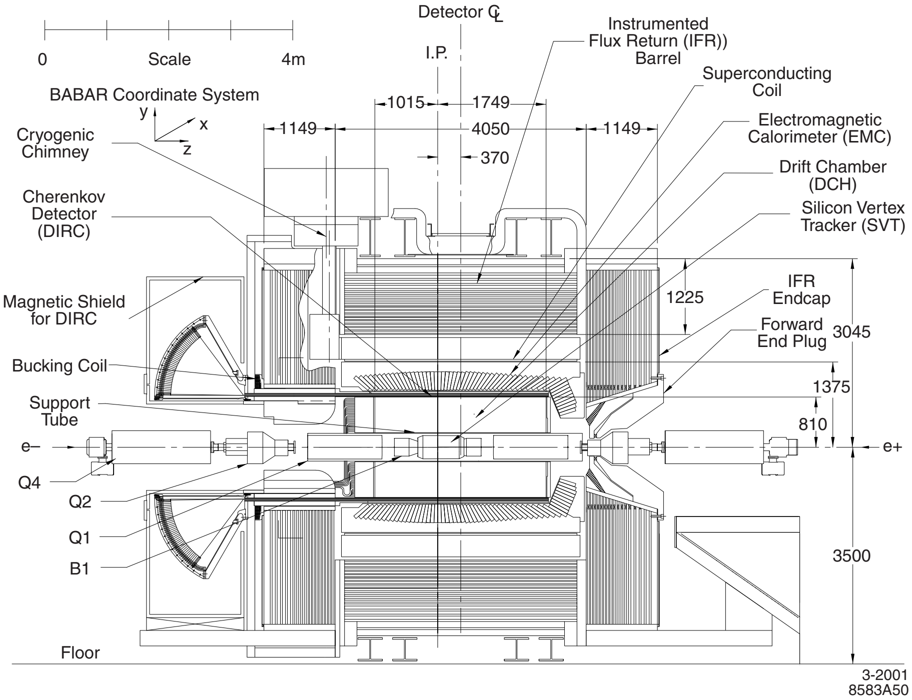
Figure 2.1: A schematic of the PEP-II elements at the *BABAR* IP.

2.4 Silicon Vertex Tracker (SVT)

The SVT, as the innermost subdetector of *BABAR*, provides high precision charged particle tracking, weak decay vertexing and determination of the IP.

Extrapolation of accurate track information from the SVT to the other subsystems is important for the refinement of their respective resolutions and for alignment. In particular the DIRC resolution is sensitive to the error on the track angle as determined from the DCH and SVT. It is therefore a requirement that the combined SVT and DCH track errors are small compared to the error on the DIRC Cherenkov angle.

The SVT is also solely responsible for providing tracking and $\mathrm{d}E/\mathrm{d}x$ information



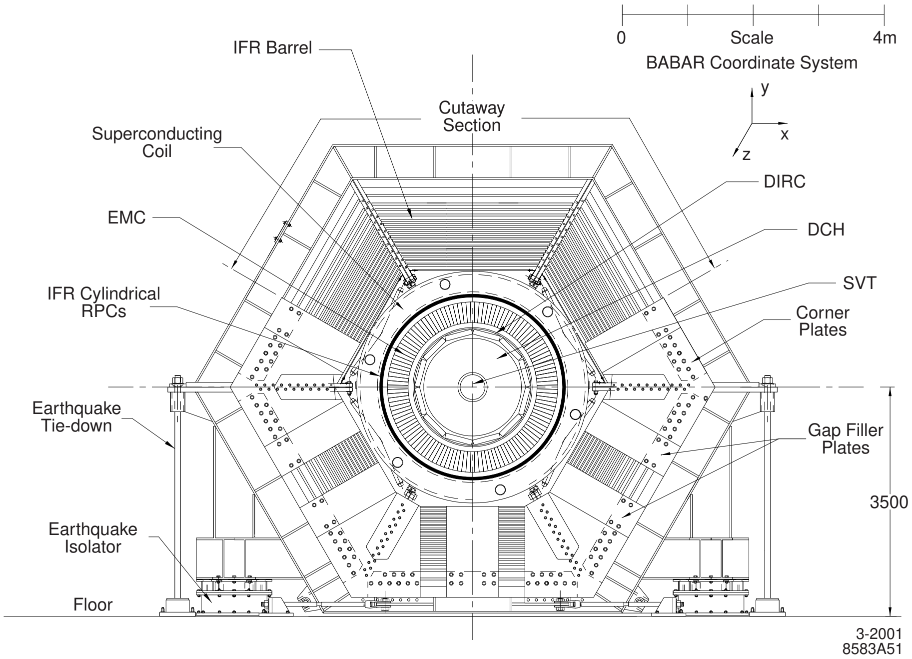


Figure 2.2: Cross sectional projections of the *BABAR* detector.

for particles with transverse momenta less than 100 MeV/c since they do not reach the DCH.

2.4.1 Design

The design requirements are primarily motivated by the need to measure accurately the decay vertices of neutral B mesons to interpret the proper time between decays for the analysis of time-dependent CP asymmetries. The high boost of the CM provides a mean B meson flight length of $\sim 250\,\mu$m in the laboratory and as the B mesons share the same lifetime, this is also the mean separation of the decay vertices. It is required that the error on the vertex separation be better than half this, thus the requirement on the z position resolution for a fully reconstructed B decay is to be $< 80\,\mu$m. The resolution in the $x - y$ plane is required to be $\sim 100\,\mu$m. Figure 2.3 shows the dimensions and structural layout of the SVT and its

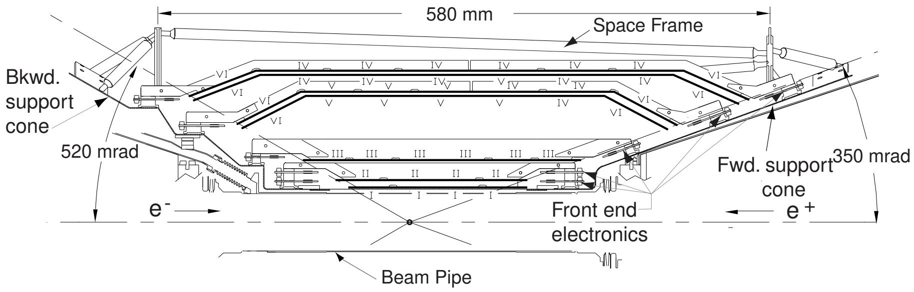


Figure 2.3: Schematic of SVT.

location with respect to the IP. The SVT is built from five layers of $300\,\mu$m thick double-sided silicon microstrip detectors, with opposing sides of strips orientated orthogonally to independently measure z and ϕ. The first three layers have six modules each, the fourth sixteen and the fifth outer layer has eighteen modules. The radial arrangement of the modules to ensure full azimuthal coverage is illustrated in Figure 2.4. Each layer is separated into a forward and backward half module for instrumentation. The assembly is supported on a Kevlar cylindrical frame. The 340 detectors are read out using 150,000 electronics channels. In total the SVT contains 0.96 m^2 of active silicon and covers 90% of the solid angle in the CM frame. Raw signals are passed through a charge-sensitive preamplifier, followed by a shaping amplifier. The strength of the signal (which is proportional to the energy loss of the

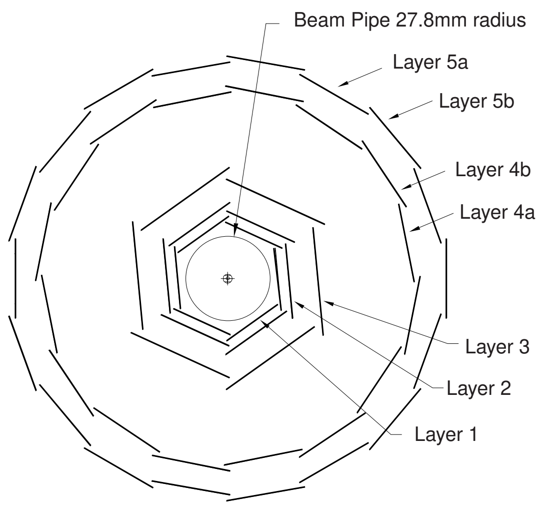


Figure 2.4: Projection along z of the SVT silicon layer arrangement.

	Layer 1	Layer2	Layer 3	Layer 4	Layer 5
Radius(mm)	32	40	54	124	142
Detectors/Layer	24	24	36	112	144
Readout Pitch (μm)					
ϕ	50	55	55	80-100	80-100
z	100	100	100	210	210
Intrinsic Resolution (μm)					
ϕ	10	10	10	10-12	10-12
z	12	12	12	25	25

Table 2.2: Performance summary of the SVT.

ionising particle) along a readout strip is compared to a threshold, and the time for which it is over this threshold (TOT) is logarithmically related to the size of the induced charge. It is this TOT information that is read out from the detector.

2.4.2 Performance and calibration

A summary of the SVT performance is given in Table 2.2, while the SVT hit resolutions as a function of z and ϕ are shown in Figure 2.5.

The efficiency of the SVT, defined as the number of associated hits to the number of tracks crossing each module, is 97%. Figure 2.5 depicts the SVT resolutions in the z and ϕ directions for each layer.

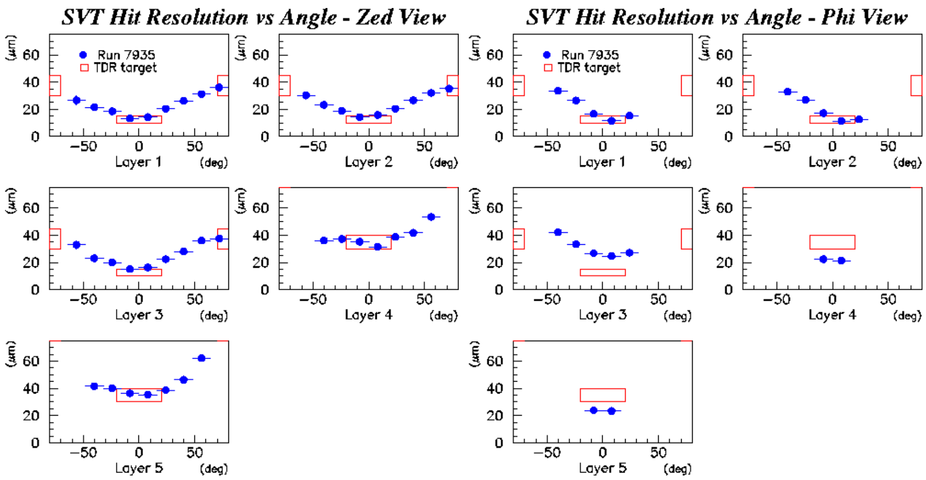


Figure 2.5: SVT hit resolutions vs z and ϕ.

Internal alignment of the SVT sensors is rather involved and is only usually performed after likely mechanical shifts, such as detector access. Cosmic rays and di-muon events are the most common tracks used for this purpose. Global alignment is performed at least once per day using track-matching algorithms between the SVT and DCH. The front end electronics are calibrated by injecting known amounts of charge into the preamplifiers and varying the TOT thresholds.

2.5 Drift Chamber (DCH)

The DCH is designed to efficiently detect and measure the momenta and track loci of charged particles with $p > 100\,\mathrm{MeV}/c$. The DCH complements the measurements of the impact parameters and directions of charged tracks seen by the SVT and provides extrapolation of tracks to the outer sub-detectors. It is used extensively for particle identification (PID) through measurements of $\mathrm{d}E/\mathrm{d}x$, particularly for tracks with momentum below $700\,\mathrm{MeV}/c$ where the DIRC is not effective. Information from the DCH is also a major input to the level 1 trigger described in Section 2.9.

The spatial resolution is required to be $\sigma(R\theta) < 140\,\mu\mathrm{m}$, with a $\mathrm{d}E/\mathrm{d}x$ error of 7%; the resolution requirement on p_t for a $1\,\mathrm{GeV}/c$ particle is $\sigma_{p_t}/p_t \simeq 0.3\%$.

2.5.1 Design

The DCH is 2.8 m long gas filled cylinder with an inner radius of 23.6 cm and an outer radius of 80.9 cm. Field and sense wires are strung longitudinally between two endplates to form forty layers of hexagonal cells. A schematic of the chamber is shown in Figure 2.6. 7104 hexagonal cells are arranged in ten superlayers each consisting of four layers of cells. The superlayers are arranged in alternating layers of axial (A) and stereo (U,V) configurations. Axial wires lie parallel to the z axis, the stereo wires are at a slight angle ($\pm 40 - 70$mrad) to z and allow determination of the longitudinal position. The layer arrangement is shown in Figure 2.7. The gas

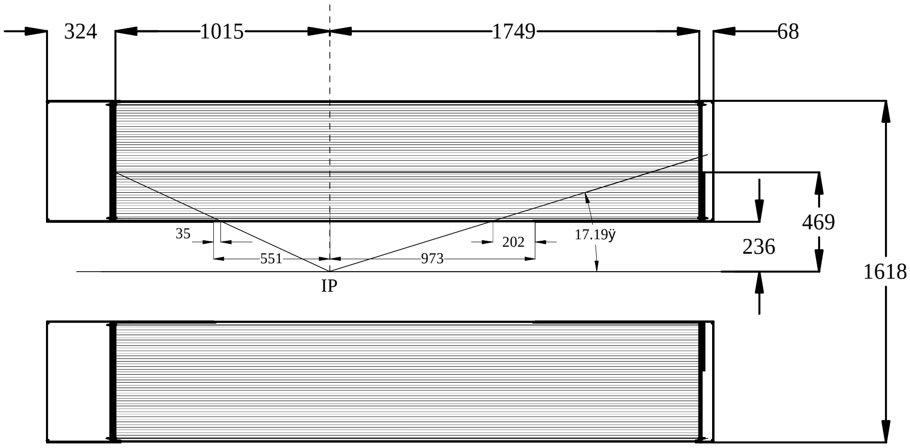


Figure 2.6: Schematic of DCH. All dimensions are in mm.

composition is helium-isobutane in the ratio 4:1. The sense wires are gold-plated tungsten-rhenium while the field wires are gold-plated aluminium and raised to an operating voltage of 1930 V. A typical minimal ionising particle will ionise $\sim$ 5 - 10 electrons/cell; the avalanche gain can be up to 5×10^4.

Minimising multiple scattering is obviously a prime concern of any tracking detector and is particularly important in *BABAR* for reducing the resolution degradation of the DIRC and EMC which lie outside it. To this end, the DCH is built of light materials while the gas mixture is helium based (the isobutane is added to absorb

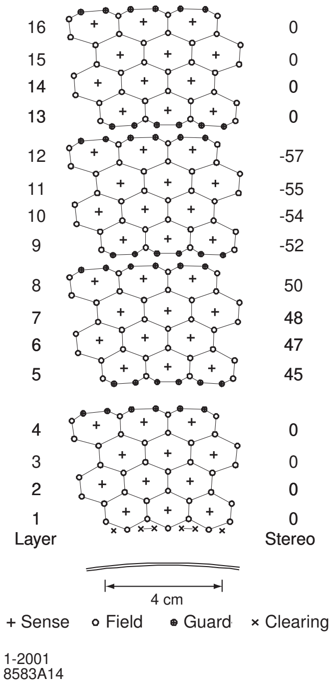


Figure 2.7: Arrangement of DCH superlayers.

photons and curtail the spread of secondary ionisation); the electronics are mounted at the rear to minimise the material in the forward region.

Measurements of the drift time from ionisation yield $x - y$ positional information about the track while the time-integrated charge deposited on each wire, being proportional to the energy loss, is used to estimate $\mathrm{d}E/\mathrm{d}x$. Shaped analogue signals from the amplifier ICs are digitised using a 6-bit 15MHz Flash ADC for $\mathrm{d}E/\mathrm{d}x$, whilst the drift time from ionisation is measured with a precision of 1ns using a 4-bit TDC.

Charged tracks are defined by five parameters $(d_0, z_0, \phi_0, \omega, \tan\lambda)$, evaluated at the point of closest approach (POCA) to the z axis. The parameters d_0 and z_0 are the distances of closest approach in the radial and z directions, ϕ_0 is the azimuthal angle of the track, $\omega = 1/p_T$ is the curvature and λ is the dip angle relative to the transverse plane.

2.5.2 Performance and calibration

The resolution on track transverse momentum is well described by the relation

$$\frac{\sigma_{p_T}}{p_T} = (0.13 \pm 0.01\%).p_T + (0.45 \pm 0.03)\% \tag{2.1}$$

as can be seen in Figure 2.8

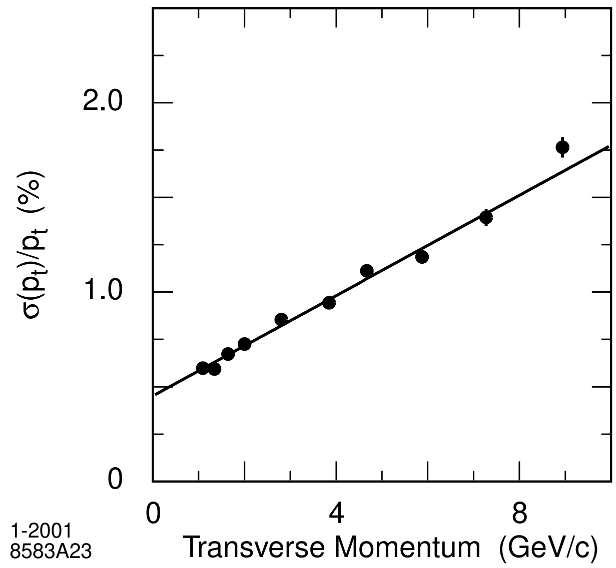


Figure 2.8: Transverse momentum resolution of tracks in the DCH. The overlaid fitted line is described by Eq.(2.1).

Figure 2.9 shows the drift chamber performance for dE/dx as a function of momentum. The $K - \pi$ overlap is visible at around 1 GeV/c. The efficiency as a function of polar angle and momentum is shown in Figure 2.10. Here two operating voltages are used for comparison, 1900 and 1960 V; currently the DCH runs at 1930V which is a compromise between detector lifetime concerns and performance.

2.6 Detector of Internally Reflecting Cherenkov Light (DIRC)

The DIRC is the primary detector for particle identification for tracks above $\sim$ 1 GeV/c since the DCH cannot distinguish kaons and pions at and above these

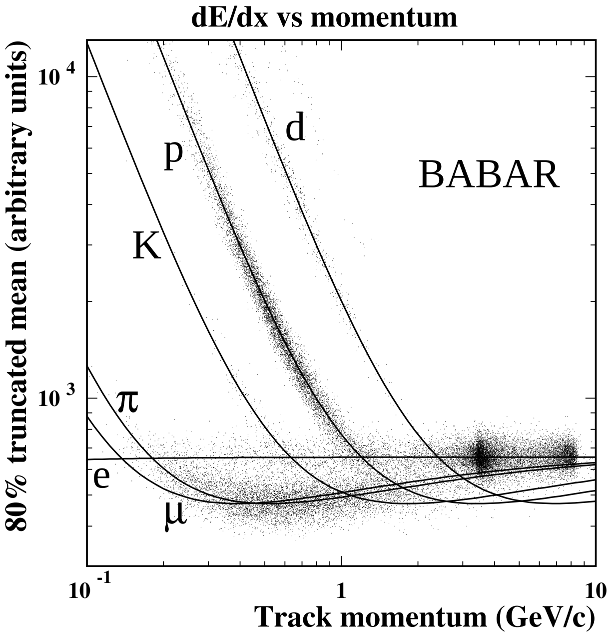


Figure 2.9: dE/dx as a function of momentum as mesured in the DCH. Overlaid are the Bethe-Bloch functions expected for various particle species.

momenta. Specifically the DIRC is charged with providing 4σ $K-\pi$ separation over the range 0.7 - 4.2 GeV/c. PID is performed by measuring the angle of Cherenkov radiation produced in the DIRC's dielectric medium of refractive index n. The velocity of the particle, obtained using the relation

$$\beta = \frac{1}{n\cos(\theta_c)}, \tag{2.2}$$

along with the momentum measurement from the track curvature, gives the mass of the particle.

2.6.1 Design

The minimisation of material in front of the calorimeter, along with compactness to keep the size and cost of the calorimeter small, are obviously critical design constraints. The solution is an assembly of 144 bars of synthetic quartz arranged symmetrically in a twelve-sided barrel. The bars are 4.9 m in length each and are made from four pieces attached end-to-end with adhesive. The radial thickness of the DIRC including supports is about 8 cm, representing only 17% of a radiation

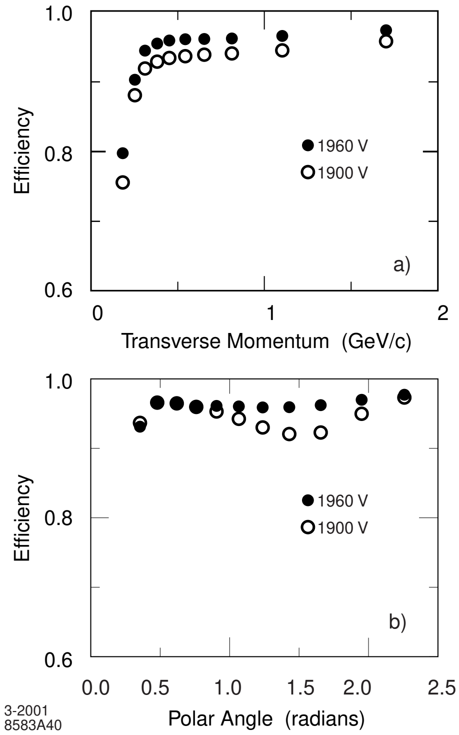


Figure 2.10: DCH efficiency as a function of polar angle and p_t. Two operating voltages are used for comparison. The DCH is currently run at 1930V.

length. Quartz is chosen for the high quality optical finishing of its transmission surfaces, its long attenuation length, low chromatic dispersion and its radiation hardness, amongst other things. The refractive index is 1.473, giving a critical angle of 42.7° and a maximum Cherenkov angle of 47.2°. The design is such that the Cherenkov angle θ_c is preserved by internal reflection in the DIRC.

Cherenkov radiation from charged particles is produced and propagated by internal reflection along the quartz bars and transmitted into a large container of 6000 litres of pure, de-ionised water called the stand-off-box (SOB), which is mounted at the rear of the detector. Water is chosen as it has a similar refractive index (1.43) to quartz, thereby lowering the refraction at the quartz termination. Mounted on the back of the SOB is an array of 10,752 photomultipler tubes (PMTs) each 29 mm in diameter. The DIRC design is unique, and the principle is illustrated in Figure 2.11. The pattern the PMTs actually see are a conic section with an opening angle of the quartz/water refracted θ_c. Specifically the DIRC measures three quantities:

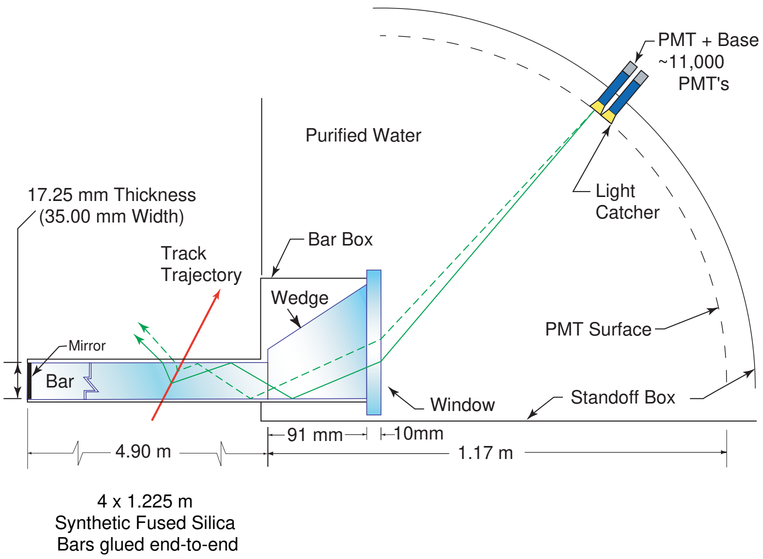


Figure 2.11: Illustration of the DIRC principle.

the propagation time information of the Cherenkov photons, the angle where the pattern appears (and hence the $x - y$ coordinate of the track at the DIRC) and the Cherenkov angle itself. Figure 2.12 taken from a di-muon event shows typical PMT patterns and highlights the importance of accurate timing information. The z coordinate of the track entry point in the DIRC and the track angle are used to predict accurately the arrival time of the Cherenkov photons in the PMTs. Hence good track parameter measurements from the DCH and SVT are essential for a narrow timing cut.

2.6.2 Performance and calibration

The single photon angular resolution is 10.2 mrad or 0.57°, the time resolution is 1.7 ns and the Cherenkov angle resolution is 2.5 mrad. The $K - \pi$ separation as a function of momentum is shown in Figure 2.13; the efficiencies as a function of momentum for identifying kaons and rejecting pions are shown in Figure 2.14.

The DIRC PMT efficiencies and timing response are calibrated using precise 1ns flashes from blue LEDs in the SOB. The calibration is performed daily as part of the global calibration.

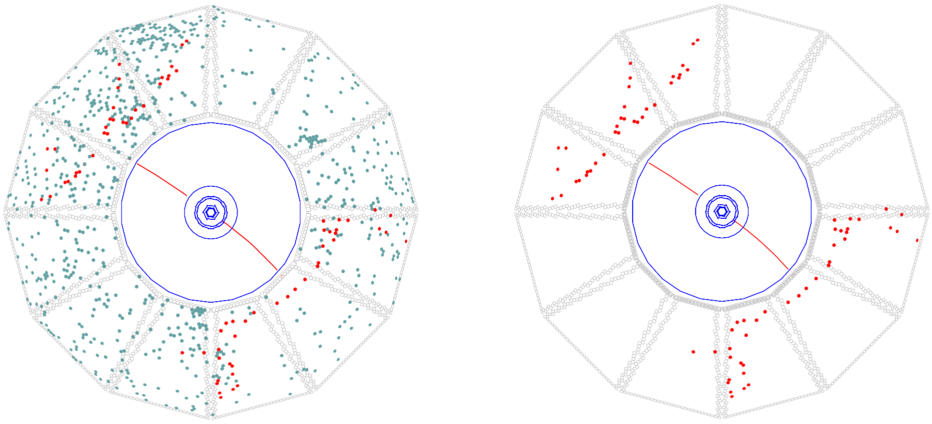


Figure 2.12: These two plots display a di-muon event with two different time cuts. In the left-hand plot, all DIRC PMTs with signals within the ±300ns trigger window are shown. In the right-hand plot, only those PMTs with signals within 8ns of the expected Cherenkov photon arrival time are displayed.

2.7 Electromagnetic Calorimeter (EMC)

The EMC is the most immodest of the *BABAR* subsystems; it was the most costly to manufacture, it draws more power for its front end electronics than any other system, makes the biggest demands on the cooling infrastructure and has the largest array of off-detector read-out modules.

It is responsible for the measurement of electromagnetic shower energies ranging from 20 MeV to 9 GeV. The lower bound is set by the requirement to reconstruct π^0 and η mesons; the upper bound is determined by Bhabhas and is equal to the energy of the HER. The system can tolerate energy depositions up to 13 GeV to allow possible future running at the $\Upsilon(5S)$. The EMC also plays a role in particle identification (particularly electrons) from measurements of the ratio of E/p for charged tracks.

2.7.1 Design

The EMC is assembled as two separate structures, namely the endcap and the barrel. The barrel is further divided into forward and backward regions for instrumentation

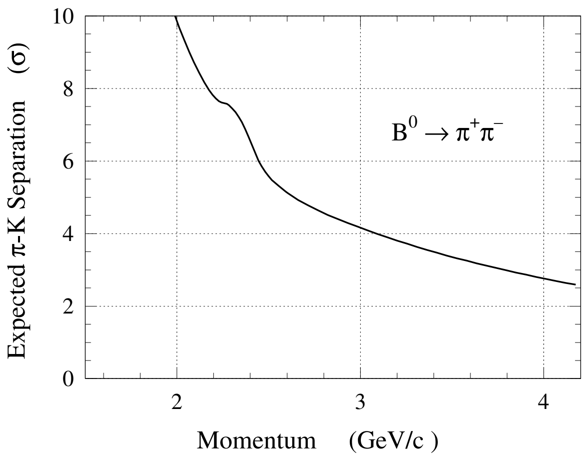


Figure 2.13: Kaon-pion separation in the DIRC as a function of momentum.

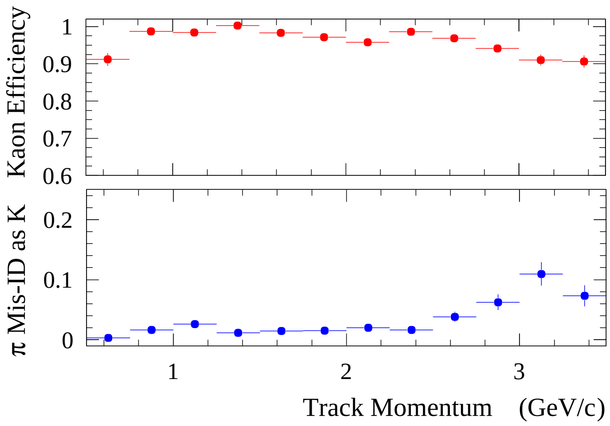


Figure 2.14: Kaon efficiency (top) and pion rejection (bottom) as a function of momentum.

purposes. Together these sections comprise 6580 caesium iodide (doped with thalium) crystals housed in carbon fibre support structures mounted on an aluminium strongback; the entire ensemble is more than 2 metres in diameter, 3 metres in length and weighs over 25 metric tonnes. The endcap is made of 820 crystals arranged in 8 concentric rings; the barrel consists of 5760 crystals divided into 48 rings of 120 crystals each. The geometrical arrangement of these crytals provides the EMC with a hermeticity in the $\Upsilon(4S)$ frame of 90%. Figure 2.15 shows a schematic of the EMC and a cut-away of the crystal assembly is illustrated in Figure 2.16.

Ceasium iodide with a thalium doping concentration of 0.1% is chosen as the detector

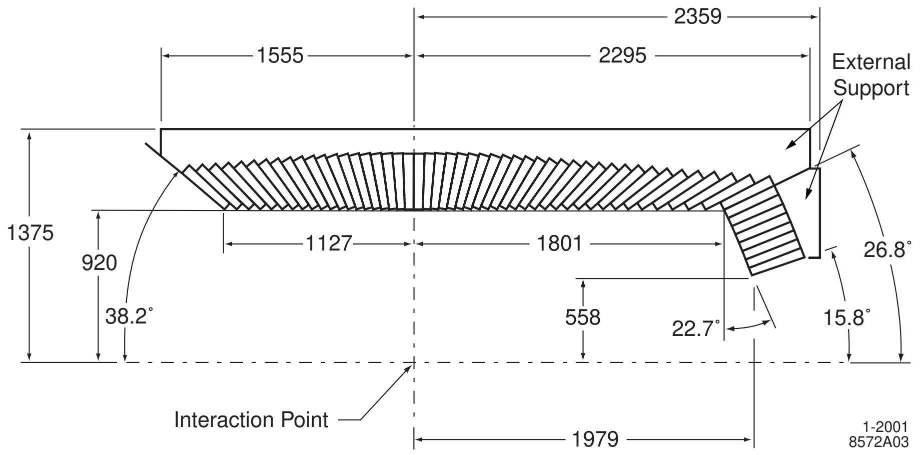


Figure 2.15: Cross section of EMC.

material because of its excellent light yield of 50,000 photons/ MeV and small Molière radius of 3.8cm. The crystals are on average $\simeq$ 17 radiation lengths deep, the nominal radiation length (X_0) for CsI being 1.85cm. The barrel and outer 5 rings of the endcap have 0.3-0.6X_0 of material between them and the IP. The inner three rings of the endcap lie in the shadow of the SVT support structure and may see up to 3.0X_0 of matter in front of them.

Attached to the back of each crystal are two solid state photodiodes sensitive to the scintillation radiation wavelength of the crystals which peaks near 560 nm. These devices are well suited to the high magnetic field environment and the use of two in parallel reduces noise in addition to offering redundancy. Analogue signals from the diodes undergo preamplification on-crystal before being transmitted via shielded flat-ribbon cables to the front end electronics (FEE) for processing.

The resolution as a function of energy for a crystal scintillation-based electromagnetic calorimeter can be described by the function

$$\frac{\sigma_E}{E} = \frac{\sigma_1}{E^{\frac{1}{4}}} \oplus \sigma_2 \tag{2.3}$$

where E is measured in GeV and $\oplus$ means to add in quadrature. The term σ_1 is due to stochastic fluctuations in shower development, photon/electron statistics and electronics noise and is dominant at low energies; the constant σ_2 term dominates

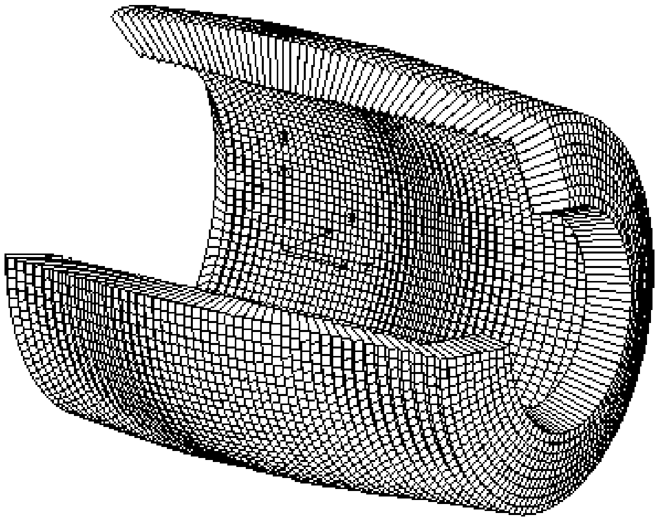


Figure 2.16: Cut-away of EMC crystal layout.

at higher energies ($> 2\,\mathrm{GeV}$) and originates largely from inhomogeneities in crystal manufacture, radiation damage to the crystals, shower leakage and preshowering from obstructing inert material.

The EMC angular resolution (both polar and azimuthal) is governed by the crystal face size and the distance from the interaction point. It takes the functional form

$$\sigma_\theta = \sigma_\phi = \frac{a}{\sqrt{E}} + b \tag{2.4}$$

where a and b are constants.

2.7.2 Performance and calibration

The low energy empirical resolution found from the radioactive-source calibration is

$$\frac{\sigma_E}{E} = (5.0 \pm 0.8)\,\%\ @\ 6.13\,\mathrm{MeV} \tag{2.5}$$

At high energies, the resolution function is determined from Bhabha scattering where the energy of the $e^{\pm}$ in a given polar angle can be calculated very accurately. It is found to be

$$\frac{\sigma_E}{E} = (1.90 \pm 0.07)\,\%\,@\,7.5\,\mathrm{GeV} \tag{2.6}$$

Figure 2.17(a) shows the measured energy resolution for the EMC derived from different physics processes as a function of photon energy. A fit to the resolution function , Eq.(2.3), below 2 GeV from these processes gives

$$\frac{\sigma_E}{E} = \frac{(2.32 \pm 0.07\%)}{E^{\frac{1}{4}}} \oplus (1.85 \pm 0.12)\%. \tag{2.7}$$

The angular resolution, derived from studies of π^0 and η decays with approximately equal photon energies, is found empirically to be

$$\sigma_\theta = \sigma_\phi = \left(\frac{(3.87 \pm 0.07)}{\sqrt{E}} + (0.00 \pm 0.04)\right)\mathrm{mrad} \tag{2.8}$$

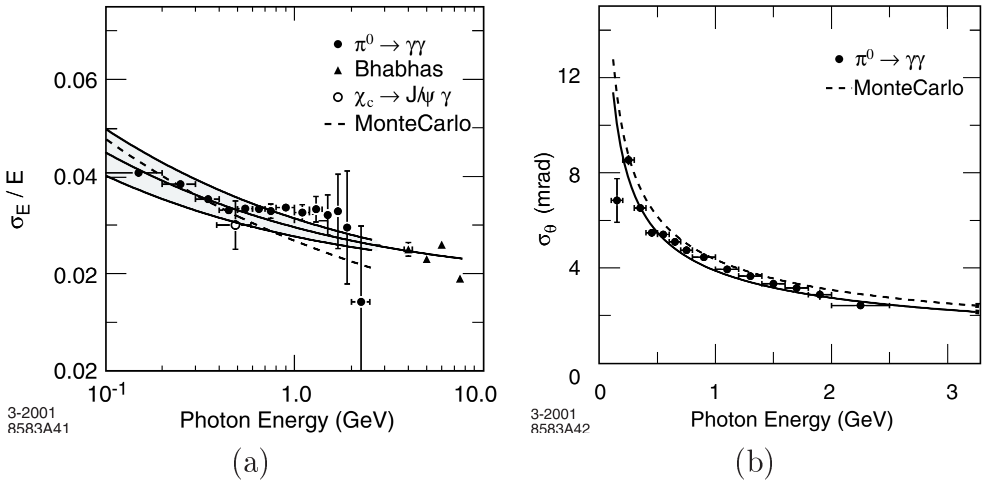


Figure 2.17: The energy resolution (a) and the angular resolution (b) of the EMC as a function of energy. In the first figure, the solid curve represents a fit to the functional form given by Eq.(2.3). In the second figure, the functional form fitted is given by Eq.(2.4).

The electronics are calibrated by injection of a known (variable) charge into the preamplifiers. The crystal light yield is determined using a 6.13 MeV radioactive γ source which is derived from irradiating FLOURINERT (PCTFE) with a neutron generator. The irradiated FLOURINERT is circulated at a rate of 125 litres/s through closed plumbing which passes in front of the crystal faces, providing photons at a rate of around 40 Hz at the crystal face. Variations in crystal to crystal light yields can be up to a factor of two. The 6.13 MeV photon also fixes the low-end energy spectrum. Bhabha processes are used to fix the upper end of the energy spectrum and these two points are used to extrapolate the crystal response over the dynamic range.

In five years of operation only two crystals have been permanently lost due to defective instrumentation.

2.8 Instrumented Flux Return (IFR)

The IFR is designed to return the flux from the superconducting magnet, to provide structural support for the *BABAR* detector and act as the principal detector for identifying muons; it also serves as a neutral hadron detector especially for the K_L^0. Muons are of importance for sin2β channels such as $B^0 \to J/\psi K_X$, both through $J/\psi \to \mu^+\mu^-$ and tagging of the other B through semi-leptonic processes. They are also critical for several other physics analyses such as τ physics, so efficient detection of these particles is essential to the physics at *BABAR*.

2.8.1 Design

The IFR is assembled as a hexagonal shaped steel support structure, with forward and backward endcaps and barrel section as illustrated in Figure 2.18. The barrel is segmented into nineteen layers of single gap resistive plate chambers (RPCs), separated by iron plates ranging from 20mm at the innermost to 100mm at the outermost in thickness. The endcap has eighteen layers of RPCs each of which is orientated orthogonal to the beam line. The RPCs are composed of 2mm thick Bakelite sheets separated by a gaseous gap of 2mm. This gap is enclosed using PVC spacers and is filled with an Argon/Freon/Isobutane gas mixture. This is all held between graphite electrodes kept at a potential difference of 8kV. When an ionising particle passes through the RPC, its ionisation trail creates a conducting path allowing a discharge of about 100pC. Outside the graphite layer and insulated from it with mylar are aluminium strips which pick up the discharge through capacitative coupling, giving a signal of around 300mV per spark. The response time is short, typically of the order 1ns or so. Opposite sided strips are arranged orthogonally giving a pitch of about 4cm in z and between 2-3cm in $r\phi$ for the barrel, and between 3-4cm in x and y for the endcap. Signals from individual IFR strips are shaped and compared to a threshold before being grouped together to form clusters. There are also two cylindrically arranged layers of RPCs between the outside of the EMC and the magnet, helping to link tracks to IFR clusters. Information from the IFR is used in the level 1 trigger, described in the next section.

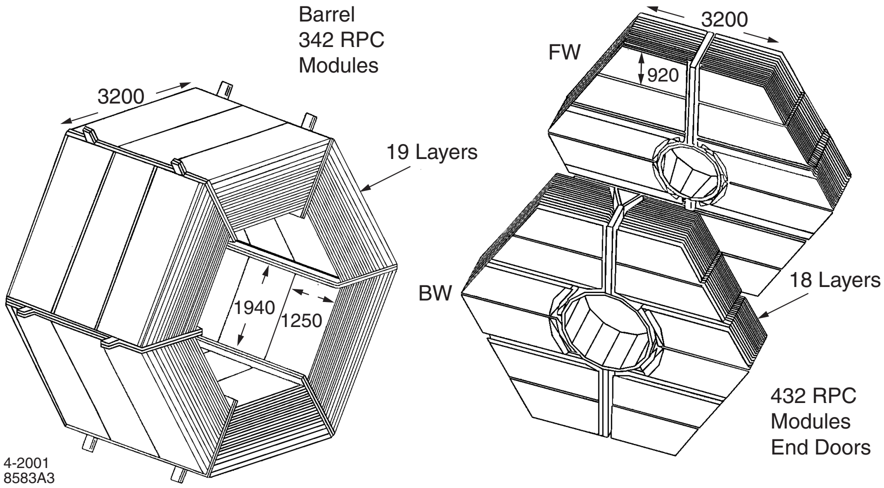
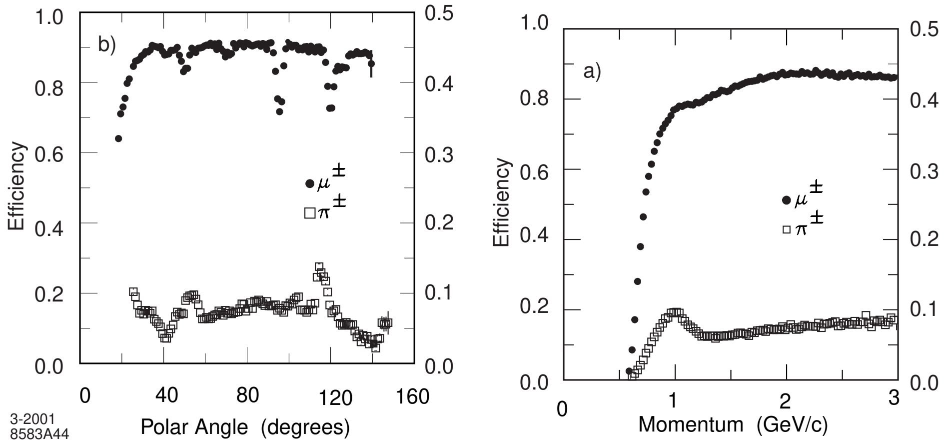


Figure 2.18: Illustration of IFR assembly. All dimensions are in mm.

2.8.2 Performance

The IFR efficiency is ascertained from normal collision data and cosmic rays and is measured weekly. The efficiency is calculated by first projecting DCH tracks through to the IFR; IFR clusters (defined as groups of adjacent hits in one of the two readout coordinates) which lie 12cm from the extrapolated track are used to form 3D clusters. A straight line fit is then performed to the resulting group of clusters and an RPC is considered efficient if it records a hit less than 10cm from the best fit line. At the start of running 75% of all active RPCs had a muon detection efficiency of above 90%. Figure 2.19 shows the muon efficiency and pion misidentification rates during Run 1.



(a) (b)

Figure 2.19: The muon efficiency and pion misidentification rates of the IFR as a function of polar angle (a) and momentum (b).

Unfortunately since the commision of the IFR, a large number of RPCs have developed problems such as large dark currents, low efficiency, and high single hit rates. The low efficiency is largely attributed to change in the Bakelite bulk resistivity, either due to overheating, running too large a charge through its volume or over-oiling of the Bakelite surfaces. Average RPC efficiencies are around 50% and current muon ID performance is already inadequate for many applications. Consequently, significant upgrades to the IFR with the replacement of the RPCs with limited streamer tubes (LSTs) were started in the 2004 shutdown period and are planned to continue during the 2005 and 2006 shutdowns.

2.9 Trigger

The purpose of the trigger system is the selection of events of interest with high efficiency. Interesting events are not just those for physics analyses, and often include what would be called 'background'. For instance Bhabha events, machine backgrounds and cosmic events all have trigger lines associated with them and are used for luminosity calculations, calibrations, diagnostics and characteristion of running conditions. The *BABAR* trigger is currently split into two stages, a hardware level 1 (L1) trigger, and a software level 3 (L3) trigger. A level 2 trigger is not currently in existence but the possibility for its future conception remains should L1 accepted event rates become too high to be handled efficiently by L3.

2.9.1 L1 trigger

The L1 trigger is the first stage in the event-rate reduction and is used to preselect events before passing them to L3 via a fast control and timing system.

The L1 is currently limited to around 3.5kHz, with upgrades that might push it up to a limit of 5kHz, beyond which readout rates are limited by things such as calorimeter feature extraction (FEX) [36]. Typically the L1 event rate output is around 2.0-2.5kHz. It is composed of the drift chamber trigger (DCT), the electromagnetic trigger (EMT), the IFR cosmic trigger (IFT) and the global tigger (GLT). The IFT is primarily used to veto cosmic ray events. The DCT and EMT construct basic objects such as track segments and EMC cluster energies which are then passed to the GLT. The GLT assembles these objects and evaluates the event based on 24 separate trigger lines. Certain combinations of these trigger lines will produce an L1 accept, whereby the whole detector is feature extracted and read out to the L3 farm for event reconstruction. The total trigger latency is fixed at 12 μs. The EMT and DCT are designed to be independently 99% efficient for $B\overline{B}$ events (such that the probability an event is not selected by either trigger is $\sim 10^{-4}$) ; the orthogonality of the two triggers also allows accurate diagnosis of their respective performances. Monte Carlo simulations indicate that the L1 trigger is 100% efficient for passing B events which fall inside the fiducial volume of the detector [35].

2.9.2 L3 trigger

The L3 trigger has as its input the output of the L1 and has access to the entire event. It runs on a dedicated 30 node batch processing farm. The L3 does not actually perform full event reconstruction, but calculates to a higher accuracy than L1 important event parameters such as track vertices, track-cluster matching, impact parameters and event topologies. Each event takes on average 8.5 ms to process. The L3 can also prescale trigger lines for event types of high luminosity, such as Bhabha events. L3 accepted events are currently being written to disk at a rate of about 300Hz, reducing the L1 rate by a factor of 10 or so. The L3 rate is principally constrained by the rate at which data can be shipped and stored. Full event reconstruction is performed offline.

Chapter 3

Event selection and continuum background fighting

3.1 Introduction

The remaining three chapters are devoted to describing in detail the analyses of the decays $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ and $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$. Since both final states are very similar, the two analyses are performed in much the same manner and as a result they are treated here in a descriptively parallel fashion. Where the analyses do sometimes take separate paths the reasons for divergence are made explicit.

This chapter begins with a description of the data and Monte Carlo (MC) samples used in this analysis before detailing the information available about the B decays at *BABAR* that is used in the identification of the signal modes and for the suppression of background. Track, π^{0} and B candidate reconstruction is described along with an overview of any corrections that must be applied to the MC to bring it in line with the data. This is followed by a description of the dominant source of background to this analysis, that of the light-quark continuum, and the efforts taken to mitigate its effect with the employment of a multivariate discriminator. The chapter then concludes with a summary of the final event selection and the efficiency tables for the selection.

3.2 Data and Monte Carlo samples

The data used in this analysis represents the full 'good-runs' data set collected since *BABAR* started in October 1999 until the end of the last run in July 2004. This complete Run 1–4 data set consists of:

- 210.6 fb^{-1} on–resonance and
- 21.6 fb^{-1} off–resonance data.

This analysis is also heavily reliant on large Monte Carlo (MC) samples which simulate the signal and background events with full detector response. The physics simulation is handled by two event generators, `EvtGen` [37] and `JetSet` 7.4 [38]. The former is used to simulate B decays to exclusive final states, the latter to model generic continuum and inclusive B decays. The detector response is fully simulated with `Geant4` [39], which not only models the physical make-up of the *BABAR* detector but also replicates the *BABAR* historical running conditions. The MC data sets used in this analysis include:

- 256 million ($230\,\mathrm{fb}^{-1}$) generic B^+B^- and 254 million ($230\,\mathrm{fb}^{-1}$) generic $B^0\overline{B}^0$ events [1].
- 240 million event continuuum sample including lighter quark and τ production ($e^+e^- \to u\overline{u}/d\overline{d}/s\overline{s}/c\overline{c}/\tau^+\tau^-$) events in the correct proportions, equivalent to $140\,\mathrm{fb}^{-1}$.
- 565k non–resonant $B^\pm \to \pi^\pm\pi^0\pi^0$ and 700k non–resonant $B^\pm \to K^\pm\pi^0\pi^0$ events.
- 1.4 million resonant $B^\pm \to \rho^\pm\pi^0$ and 1.2 million resonant $B^\pm \to K^{*\pm}\pi^0$ ($K^{*\pm} \to K^\pm\pi^0$) events.

[1] Signal events and charmless B–related backgrounds are removed 'by hand'. Every charmless background mode removed from the generic B samples was studied individually.

- More than 30 other individual charmless B decay modes with sample sizes ranging typically between 100k to 1 million events depending on the mode.

Sometimes it is necessary to correct the simulation when it poorly models the real data. These corrections are discussed below where appropriate.

3.3 Event selection

The information used in this analysis for the selection of signal-like events comes generally in the form of kinematic, topological and probabilistic-based variables that are broadly used in this analysis in three possible ways:

i For discrete selection or 'cutting'. An event or reconstructed signal-candidate can be rejected if a variable takes a value above or below some required threshold.

ii As an input to a multivariate discriminator. Rather than cut on a variable, it can be used in conjunction with others to map a multidimensional space. A discrimination surface is then formed in this manifold which is used to identify events in different classes.

iii As an input to a maximum likelihood fit. In this case the event is weighted on its likelihood for belonging to a particular class based on the value it takes for the variable.

What follows is a description of the variables used and the justification for their inclusion in the selection process.

3.3.1 *B* meson kinematics at *BABAR*

In the CM frame the energy of each B meson is, to a very good approximation, equal to the energy of one of the beams E^*_{beam}. The accuracy of the beam energy measurement determined by PEP-II is superior to the resolution on the mass of B candidates as determined from the reconstruction of the final state. This feature is exploited at *BABAR* with the construction of two variables, the "energy difference" ΔE and the "beam-energy substituted mass" m_{ES}. These are defined respectively as

$$\Delta E = E_B^* - E_{\text{beam}}^* \quad \text{and} \quad m_{ES} = \sqrt{E_{\text{beam}}^{*2} - \mathbf{p}_B^{*2}}, \tag{3.1}$$

where E_B^* and $\mathbf{p}_B^*$ are the reconstructed CM energy and momentum of the B candidate. These variables are both Lorentz invariant and are in most cases uncorrelated when the B candidate is properly and completely reconstructed. For a general final state which is fully reconstructed, the distributions of ΔE and m_{ES} are approximately Gaussian centred on $0\,\text{GeV}$ and $m_B = 5.279\,\text{GeV}/c^2$ respectively, though they often tend to have tails on the low side due to detector resolution effects. The resolution in m_{ES} is dominated by the beam-energy resolution and is almost an order of magnitude smaller than the resolution obtained by reconstructing the invariant mass from the detector information alone. It is also relatively insensitive to the nature of the final state of the B decay, in contrast to ΔE which is dominated by the detector resolution and is very much final state dependent. In this analysis, due to the final state having two π^0 mesons, the ΔE distribution is particularly broad and has a pronounced tail. This is a consequence of reconstructing photons in the calorimeter which can tend to underestimate the true energy of the particle due to the effects of shower leakage, light yield leakage and preshowering from material in front of the EMC.

ΔE also has a dependence on the mass hypotheses of the particles involved in the final state, and therefore has a different value for a given B candidate depending on whether one assumes the charged particle associated with the track is a kaon or a pion. The distribution of MC events in the (ΔE,m_{ES}) plane for true $B^\pm \to \rho^\pm\pi^0$ signal events are shown in Figure 3.1, for each mass hypothesis. Notice that the assumption of the kaon mass hypothesis for the pion in effect adds more energy to the event thereby shifting the distribution to higher values of ΔE. For the $B^\pm \to \rho^\pm\pi^0$ analysis, all tracks are assumed to be pions whereas for the $B^\pm \to K^{*\pm}\pi^0$

analysis they are assigned a kaon mass. This naturally ensures the respective signal distributions are centred properly. ΔE is a powerful tool for discriminating against

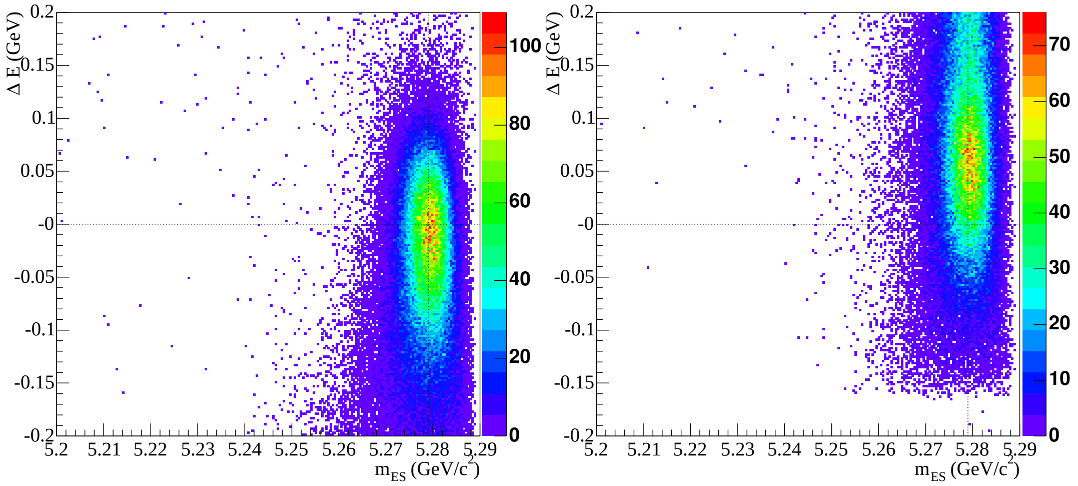


Figure 3.1: Distribution of MC events in the (ΔE,m_{ES}) plane for correctly reconstructed $B^{\pm} \to \rho^{\pm}\pi^{0}$ signal events. The left hand plot shows the distibution assuming the (correct) pion hypothesis for the track, the right shows how the distribution changes when the kaon mass is assumed.

B-related backgrounds. For example the decays $B^{\pm} \to \pi^{\pm}\pi^{0}$ and $B^{0} \to \rho^{\pm}\rho^{\mp}$ both share final states which are similar to the final state for $B^{\pm} \to \rho^{\pm}\pi^{0}$. In the former case one would typically include a soft π^{0} from the rest-of-event (ROE) to reconstruct the B, in the latter the B is misreconstructed by leaving out a slow track from the final state. In each case the reconstruction will display distributions which are peaked around the B mass in m_{ES} (as for signal), but they will be displaced to higher (lower) values in ΔE depending on whether one reconstructs $B^{\pm} \to \pi^{\pm}\pi^{0}$ ($B^{0} \to \rho^{\pm}\rho^{\mp}$). This is illustrated in Figure 3.2.

For continuum, the distributions in ΔE and m_{ES} do not peak. The continuum distribution is relatively flat in ΔE, though in m_{ES} it tends to rise slowly before falling sharply at the kinematic limit of 5.29 GeV/c^2.

In this analysis the ΔE and m_{ES} variables are used extensively, both to preselect events and as inputs to the maximum likelihood fit.

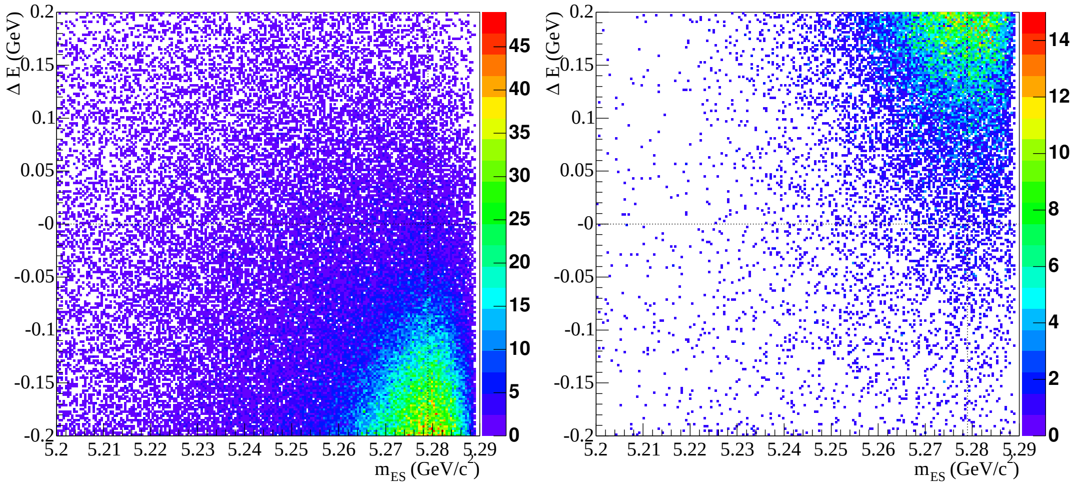


Figure 3.2: Distribution of MC events in the $(\Delta E, m_{ES})$ plane for $B^0 \to \rho^\pm\rho^\mp$ (left) and $B^\pm \to \pi^\pm\pi^0$ (right) events reconstructed as $B^\pm \to \rho^\pm\pi^0$.

3.3.2 Event topology of B decays

B mesons produced as pairs in the decays of the $\Upsilon(4S)$ have small values of momentum in the CM frame, typically around $\sim 340\,\mathrm{MeV}/c$. This, along with the fact that there is no particular preferred direction for the decays gives B events a spherical or isotropic topology. This is in contrast to events of light quark continuum production $(e^+e^- \to q\bar{q})$. The primary quarks are produced with large momenta and the resulting hadronisation process produces two highly directional back-to-back jets. A variety of variables which take advantage of this topology are used in this analysis to reduce the continuum background and are described below.

Sphericity

The sphericity axis of a system of particles is determined by the principal eigenvector of the sphericity tensor [40], defined as

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i \mathbf{p}_i^2} \tag{3.2}$$

where $\alpha, \beta = 1, 2, 3$ correspond to the x, y, z components respectively. The sphericity, S, is defined as

$$S = \frac{3}{2}(\lambda_2 + \lambda_3) \tag{3.3}$$

where λ_2 and λ_3 are the secondary and tertiary eigenvalues of the sphericity tensor respectively. For an isotropic event, all three eigenvalues should be equal to 1/3, giving S=1 [20].

By taking the scalar product of the sphericity axis of the rest of event with the sphericity axis of the B, one is able to form the variable $\cos\theta^B_{Sph}$, with θ^B_{Sph} being the angle between the two axes. For signal events the sphericity axis of the signal B should be uncorrelated with that of the ROE, which is essentially the decay of the other B. For continuum events, the two-jet-like nature ensures any B candidates formed will tend to have a sphericity axis highly correlated with that from the ROE. This is due to the decay products all being part of the same fragmentation process and all tending to have momentum along a common axis. Consequently, continuum events tend to have values of $|\cos\theta^B_{Sph}|$ close to one since the two axes tend to be more parallel, whereas true signal events are more uniform in $\cos\theta^B_{Sph}$. The distributions of $|\cos\theta^B_{Sph}|$ for signal $B^{\pm} \to K^{*\pm}\pi^0$ and continuum MC are shown in Figure 3.3 (the distribution for signal $B^{\pm} \to \rho^{\pm}\pi^0$ events is nearly identical to $B^{\pm} \to K^{*\pm}\pi^0$).

The use of the sphericity values and axis direction as input variables in a multivariate discriminant is investigated and detailed in Section 3.5.

Thrust

The thrust axis, $\hat{\mathbf{T}}$, of a system of particles is defined as the direction which maximises the sum of the longitudinal momenta. The thrust is related to this direction by

$$T = \frac{\sum_i |\hat{\mathbf{T}}.\mathbf{p}_i|}{\sum_i |\mathbf{p}_i|} \tag{3.4}$$

The allowed range of T is [0.5,1], where $T = 1$ corresponds to a jet–like event, and $T = 0.5$ corresponds to an isotropic event. As in the case for sphericity, the thrust

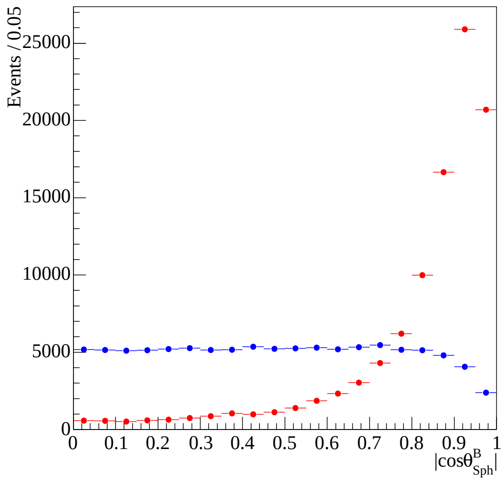


Figure 3.3: Distribution of $|\cos\theta^B_{Sph}|$ for signal $B^\pm \to K^{*\pm}\pi^0$ MC events (blue) and continuum MC (red). The fall-off at $|\cos\theta^B_{Sph}| \simeq 0.8$ for the signal distribution is due to a cut made in $|\cos\theta^B_{Thrust}|$ at the preselection level.

axis of the ROE for a true $b\bar{b}$ event should be uncorrelated with the thrust axis of the signal B. The cosine of the angle between these two vectors, $\cos\theta^B_{Thrust}$, is used at preselection to reject light–quark continuum events.

The thrust and $\cos\theta^B_{Thrust}$ were both investigated for use as input variables in a multivariate discriminant, detailed in Section 3.5.

Aplanarity

The aplanarity A is related to the smallest eigenvalue of the sphericity tensor by

$$A = \frac{3}{2}\lambda_3. \tag{3.5}$$

It is essentially a measure of the transverse component of momentum out of the plane defined by the two largest sphericity eigenvectors (i.e. the plane normal to the smallest eigenvector). The aplanarity of the rest of the event is investigated for inclusion as an input in the multivariate discriminant.

Fox-Wolfram moments

The Fox-Wolfram moments [41] are defined as

$$F_l = \sum_{i,j} \frac{|\mathbf{p}_i||\mathbf{p}_j|}{E_{vis}^2} P_l(\cos\theta_{ij}) \tag{3.6}$$

where P_l are the Legendre polynomials, p_i is the momentum for particle i, θ_{ij} is the opening angle between the i^{th} and j^{th} particle and E_{vis} is the total energy visible in the event.

These moments are commonly used as ratios, such as R2, the ratio of the second to zeroth moment. For an approximately spherical event, R2 takes a value close to zero, while for a two-jet-like event it takes a value closer to one. The Fox-Wolfram moments are investigated for inclusion as an input in the multivariate discriminant.

3.3.3 Event preselection

BABAR events have a large set of 'tags' associated with them which classify the event based on certain distinguishing features such as the number of charged tracks, total transverse momentum, the position of the primary vertex, *etc.* One such tag is the charmless `CPi0Pi0` three-body tag, which is broken down into several separate parts:

- The total event energy must be less than 20 GeV.
- Candidate B mesons are reconstructed from a charged track and two π^0 candidates. The details of the reconstruction are given in Section 3.4.
- Reconstructed B candidates must satisfy the requirement $|\Delta E| < 0.45\,\mathrm{GeV}$ assuming a pion mass hypothesis for the track.
- Candidates must pass a cut on the angle between the thrust axis of the B candidate and the thrust axis of the ROE of $|\cos\theta^B_{Thrust}| < 0.95$.
- There must be at least one charged track in the ROE.

The output of the `CPi0Pi0` tagging module is a list of pointers to accepted B candidates that are produced and stored in a central database. Events which have a candidate that pass these requirements are said to carry the `CPi0Pi0` tag-bit. Preselecting on this tag-bit is both fast and efficient and is performed before analysing the event in any further detail. Once an event is accepted, more information about the quality of the B candidate(s) is accessed and evaluated for further discrimination.

In addition to preselecting on the `CPi0Pi0` tag-bit, a cut on $|\cos\theta^{B}_{Sph}| < 0.9$ is made.

3.4 Candidate reconstruction

B candidates are constructed as composite objects from a charged track and two π^0 candidates. Construction is straightforward; the four-momenta of the decay products or 'daughters' are simply added to form the B meson.

The particles which enter as inputs to the reconstruction software are stored, for each event, in lists based on their species type. The lists exist in various sub-classes which define collections of differing purity. The lists used in this analysis are named `Pi0Loose` and `GoodTracksLoose` and their requirements are described below.

Selection criteria are applied to the photons comprising the π^0 candidates, the π^0 candidates themselves, the tracks, the invariant mass of the *lighter* pair of the track and π^0 (discussed below), and finally the B candidate itself.

3.4.1 π^0 reconstruction

Individual photon candidates are identified with single-bump clusters in the EMC that have no associated track. To reconstruct π^0 mesons, the four-momenta of pairs of distinct photon candidates (i.e. a single photon cannot be combined with itself) are added where the momentum vector of each photon is calculated assuming production at the IP. There are two types of π^0 candidate that can in principle

π^0 selection		
E_γ	$>$	0.03 GeV
LAT_γ	$<$	0.8
E_{π^0}	$>$	0.2 GeV
0.10	$< M_{\pi^0} <$	$0.16\,\mathrm{GeV}/c^2$

Table 3.1: Summary of π^0 selection requirements. The composite π^0 masses are fitted assuming production at the primary vertex.

enter into the selection: resolved or 'composite' candidates, and 'merged' candidates. Merged candidates have high momentum such that their daughter photons are nearly collinear and are not resolved into separate clusters in the EMC. For this analysis, merged π^0 candidates are not used, primarily because of the poor knowledge about the data:MC efficiency differences. The selection requirements placed on the photons and resulting composite π^0 candidates are summarised in Table 3.1. In the table LAT_γ is the lateral moment of the shower shape. This is defined to be [20]

$$LAT = \frac{\sum_{i=3}^{N} E_i r_i^2}{\sum_{i=3}^{N} E_i r_i^2 + E_1 r_0^2 + E_2 r_0^2} \tag{3.7}$$

where there are N crystals in the cluster and E_i is the energy deposited in the i^{th} crystal with the numbering defined hierarchically with E_1 being the most energetic deposition. r_i is the distance of the centre of the i^{th} crystal face from the centre of the shower, and r_0 is the length of the side of an average crystal face ($\simeq$ 5cm). This variable is designed to discriminate between electromagnetic and hadronic showers, the former of which are usually confined to just two or three crystals and thereby have smaller values of LAT.

A plot of the distributions for the π^0 invariant mass is shown in Figure 3.4 for π^0 candidates in both data and non-resonant $B^\pm \to \pi^\pm\pi^0\pi^0$ simulated events which pass the preselection.

Neutrals correction

There exist differences between data and MC simulation in the resolutions and efficiencies of neutral objects. The MC simulation produces distributions of neutral

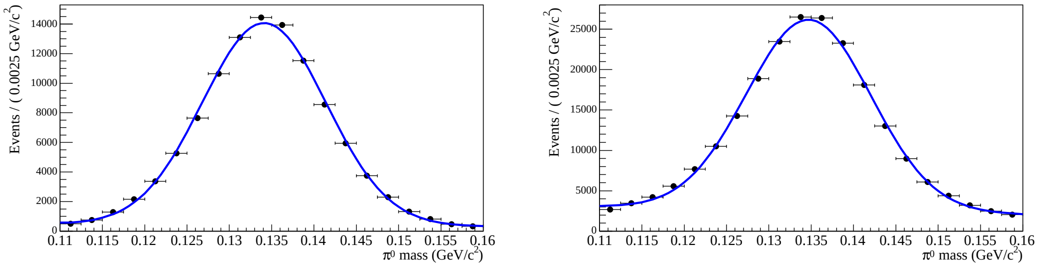


Figure 3.4: Distribution of π^0 invariant mass for non-resonant $\pi^\pm\pi^0\pi^0$ simulated events (left) and for data events (right) which pass the preselection.

quantities which are often too sharp, too clean and sometimes off-center when compared to the equivalent distribution in data. To correct for this, standard recipes have been provided within the experiment to correct the MC events and calibrate the data such that they agree [42]. Before π^0 reconstruction, changes to the photon energies and resolutions are made by way of 'shifting' and 'smearing' corrections which are derived from two methods. These methods both involve reconstructing π^0 candidates appearing in a single sample of τ 1-on-1 events. The first is through the use of a sample of symmetric π^0 decays ($\pi^0 \to \gamma\gamma$), the second through use of single-γ converted/Dalitz decays ($\pi^0 \to \gamma\gamma(e^+e^-)$). In each case, the π^0 samples are split into bins of energy before fitting for the width and mean of the $\gamma\gamma$ invariant mass distribution. These are then fed into the correction

$$\left[\frac{\Delta m_{\pi^0}}{m_{\pi^0}}\right]^2_{Data} - \left[\frac{\Delta m_{\pi^0}}{m_{\pi^0}}\right]^2_{MC} = \left[\frac{\Delta E_{average}}{E_{average}}\right]^2_{Data} - \left[\frac{\Delta E_{average}}{E_{average}}\right]^2_{MC} \tag{3.8}$$

The differences between data and MC simulation for the widths and means of the $\gamma\gamma$ invariant mass are shown in Figure 3.5.

After π^0 reconstruction, further corrections must be made to account for the overall difference in π^0 reconstruction efficiency between simulation and data. This correction is also derived from studies of τ decays, using ratios of the data:MC efficiency differences in the decays $\tau \to \rho\nu$ and $\tau \to \pi\nu$ [43]. There are systematic uncertainties related to the implementation of this correction, which have their origins in the

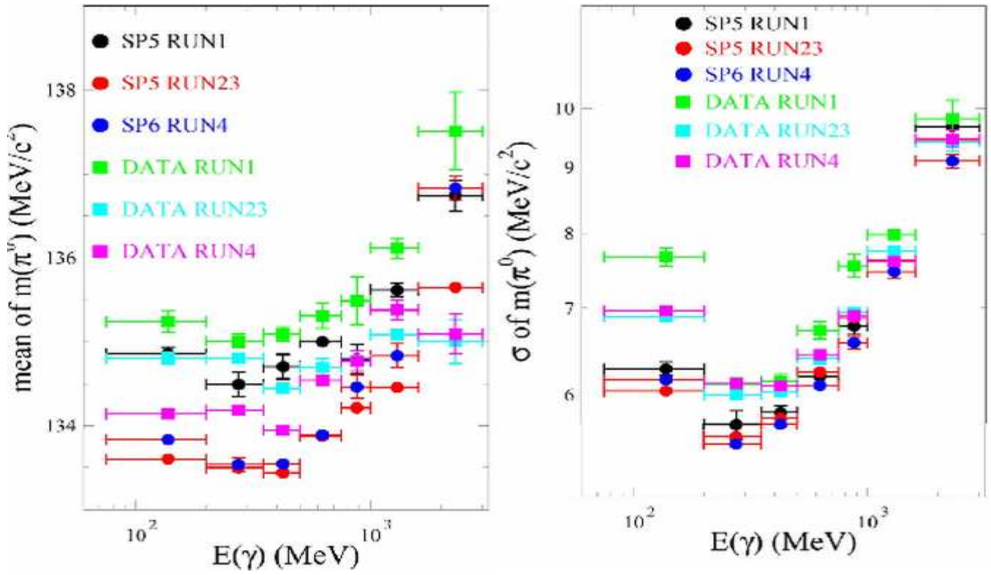


Figure 3.5: Data-MC comparison for the widths and means which are used in the γ energy smearing and shifting corrections. The letters 'SP' denote the MC simulation, with the numeral identifying the version of the generation and processing software used.

poor modelling of hadronic interactions in the EMC and the photon background. The size of the systematic error is discussed in Chapter 5.

3.4.2 Track identification

Preselection

Tracks are reconstructed from hits in the SVT and DCH. They undergo a simple preselection before being identified for species type. The preselection requirements for all tracks are summarised in Table 3.2 and are based on the track momentum magnitude (p), transverse momentum (p_T), distance-of-closest-approach (DOCA) to the primary vertex and the number of hits in the DCH ($N_{\mathrm{DCH}}^{\mathrm{Hits}}$). As with the neutrals, there are corrections that must be applied to resolve data-MC simulation differences.

Track preselection		
0.410	$\leq \theta \leq$	2.54
p_T	$>$	0.1 GeV
$0.1\,\mathrm{GeV}/c$	$\leq p \leq$	$10.0\,\mathrm{GeV}/c$
$N_{\mathrm{DCH}}^{\mathrm{Hits}}$	$\geq$	12
DOCA_{xy}	$<$	1.5cm
-10cm	$< \mathrm{DOCA}_z <$	10cm

Table 3.2: Summary of track preselection requirements.

Particle identification (PID)

There are software tools developed at *BABAR* for the identification of kaons, electrons, protons, muons and pions. The `SMSKaonSelector` [44] uses dE/dx and θ_c information from the SVT, DCH and DIRC to form a likelihood based on a classification for each track. The likelihood is formed from three 'pull' probability density functions (PDFs), one for the Cherenkov angle in the DIRC, one for dE/dx in the SVT and one for dE/dx in the DCH. For the dE/dx cases, the momentum as measured from the track locus and the Bethe-Bloch function [15], may be used to calculate the expected values of dE/dx in both the DCH and the SVT. The PDFs are formed from the Gaussian pull distribution of $[dE/dx_{Exp} - dE/dx_{Measured}]/\sigma$ and return a different likelihood depending on the particle species assumed for the expected distribution. A similar pull is defined for the Cherenkov angle. The selector can be run in several different modes depending on the level of efficiency or purity one desires.

The selector is used in the $B^{\pm} \rightarrow K^{*\pm}\pi^0$ analysis in 'Loose' mode for the selection of kaons, which keeps the kaon misidentification rate below 7% up to a momentum of $4\,\mathrm{GeV}/c$ [44]. The data and MC efficiencies as a function of momentum for the Loose selection are shown in Figure 3.6.

For pion selection, it is required that the track fails the 'Tight' proton selection [45], the 'Tight' kaon selection and the 'Not–a–pion' kaon selection.

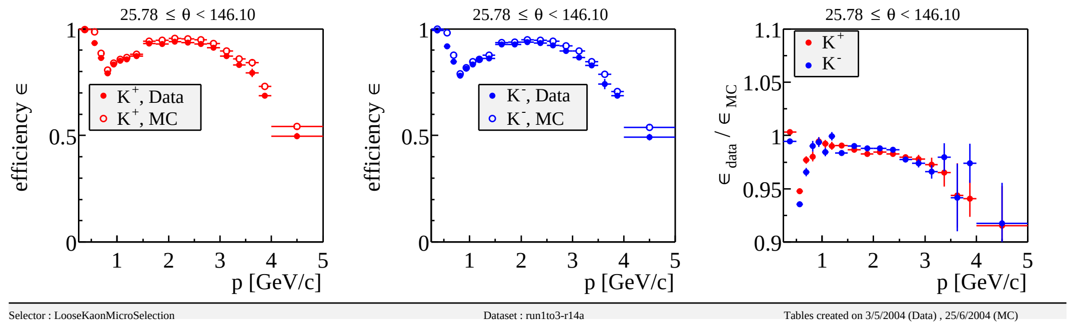


Figure 3.6: Performance of kaon 'Loose' selection as a function of momentum for positive tracks (left) and negative tracks (middle). The overall data:MC comparison is shown on the right.

Track and PID corrections

Similarly to the neutrals, corrections have been implemented for tracks to bring the MC simulation in line with the data. The corrections exist as ratios of data:MC efficiency for tracks of differing species type in bins of momentum, θ and ϕ. The corrections are made from control samples of fully reconstructed decays such as $D^{*+} \to \pi^+ D^0(\to K^- \pi^+)$. The implementation of the correction is straightforward. First MC truth information is used to identify the species of the track. The PID tables are then used to determine the selection probability for this track. A random number is then generated and the track is either accepted or rejected based on whether it is less than or greater than the probability. Applying this correction has a systematic error associated with the uncertainty on this probability from the sample statistics used to create it. This systematic error is discussed in Chapter 5.

3.4.3 K^* and ρ mass selection

To further improve the signal/background ratio, a selection cut is applied on the invariant mass $m(h^{\pm}\pi^0)$ of the resonance candidate.

The choice of mass cuts is different for each analysis. In the case of the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, the cuts were kept fairly tight. As one opens up the cut on the high mass side, one allows events from higher $K^{**\pm}$ resonances to enter into the sample through their tails. The amounts of higher $K^{**\pm}$ pollution are not known, as there have been no measurements to date of the branching fractions of $B^{\pm} \to K^{**\pm}\pi^0$. It is important to try and exclude these events as their final states are extremely similar in the discriminating variables used in the fit; except for the $K^{\pm}\pi^0$ invariant mass these modes peak in all the same places as the $B^{\pm} \to K^{*\pm}\pi^0$ signal. They can thus contribute an excess of signal events as it is difficult to identify them as being from higher $K^{**\pm}$ events. The full treatment of the higher $K^{**\pm}$ problem is left until the next chapter, but there is motivation for the mass cut to exclude these higher resonant modes as much as possible. An upper bound cut of $m_{K^{\pm}\pi^0} < 1.0\,\mathrm{GeV}/c^2$ is applied. On the low side, below about $0.8\,\mathrm{GeV}/c^2$, the invariant $K^{\pm}\pi^0$ mass spectrum is dominated by continuum events. A cut of $m_{K^{\pm}\pi^0} > 0.8\,\mathrm{GeV}/c^2$ removes only 4% of the truth-matched signal events but removes about 55% of the continuum events that survive the upper bound cut. A plot of the $K^{\pm}\pi^0$ invariant mass spectrum for truth-matched signal $K^{*\pm}\pi^0$ is shown in Figure 3.7.

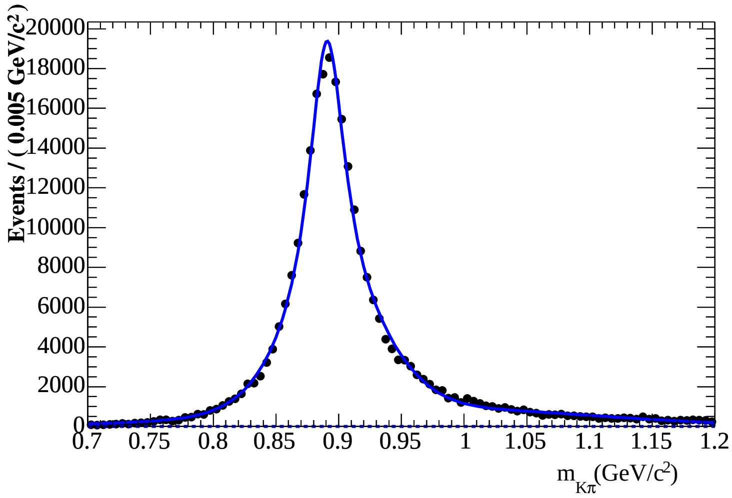


Figure 3.7: $K^{\pm}\pi^0$ invariant mass for truth-matched signal. Overlaid is a fitted Breit-Wigner lineshape, which describes the distribution well.

In the case of the $B^{\pm} \to \rho^{\pm}\pi^0$, the analysis is not approached as a search but rather to improve upon an already existing *BABAR* asymmetry measurement for this mode. An important factor to consider then is the fraction of misreconstructed signal events that have the wrong charge. This form of misreconstruction is called 'mistag'. The mistag reconstruction has the effect of washing out the observation of any asymmetry. Consequently the motivation for the mass cut is to be as exclusive as is reasonable for this self-cross-feed component whilst keeping a good proportion of truth-matched events. A variety of mass cuts were investigated with cuts both symmetrically and asymmetrically either side of the nominal ρ mass of 770 MeV/c^2. These are summarised in Table 3.3. It is important to appreciate the qualitative difference between the mistag fraction and the truth-matched efficiency as they are not directly comparable quantities. The truth-matched efficiency is the number of candidates after all selections that are correctly reconstructed compared to the number of events that were generated, whilst the mistag fraction is the fraction of events that pass all selections that carry the wrong charge and so must be incorrectly reconstructed. Choosing the range based on the lowest mistag fraction may be at the expense of the truth-matched efficiency and vice-versa, and it is clear that some balance must be found between the two. The ratio of truth-matched efficiency to mistag fraction was considered a reasonable quantity by which to choose the mass range.

An asymmetric mass cut of $0.55 < m_{\pi\pi^0} < 0.95\,\mathrm{GeV}/c^2$ was chosen for having the best ratio of truth-match efficiency to mistag fraction. The two numbers that appear in Table 3.3 for this selection for truth-matched efficiency and mistag fraction are not final however, as there are two further selections yet to be made which affect them. These selections are discussed below.

One might reasonably think that there will be two choices available for the resonance candidate as one could pair the track with either π^0. For the $K^{*\pm}\pi^0$ analysis, there is actually no choice in resonance pair at all - it must always be the lighter pair. This is because there is *no ambiguity* in selecting the $K^{*\pm}$ candidate as there is no overlap between the choice of pairs for the resonance in question due to both the kinematic boundaries of the plot and a reasonably chosen mass range. One would have to make a cut on the $K^{*\pm}$ mass in excess of 700 MeV/c^2 from its nominal value (more than 14 times its nominal width [15]) for there to be an issue of ambiguity. Thus there is only one invariant mass pair that can be reasonably formed in the resonance region

Summary of ρ mass cut investigations			
Mass cut range (GeV/c^2)	Truth-match efficiency ε (%)	Mistag fraction f (%)	Ratio of $\varepsilon : f$
0.37 - 1.17	9.552	18.65	0.51
0.47 - 1.07	9.575	16.22	0.59
0.50 - 1.00	9.524	14.46	0.66
0.52 - 1.02	9.473	14.79	0.64
0.55 - 0.95	*9.286*	*12.90*	*0.72*
0.57 - 0.97	9.231	13.28	0.70
0.60 - 1.00	9.066	13.94	0.65
0.67 - 0.87	7.458	10.31	0.72

Table 3.3: Summary of ρ mass cut investigations. The chosen mass cut is indicated in red italic, giving the best ratio of truth-matched efficiency to mistag fraction.

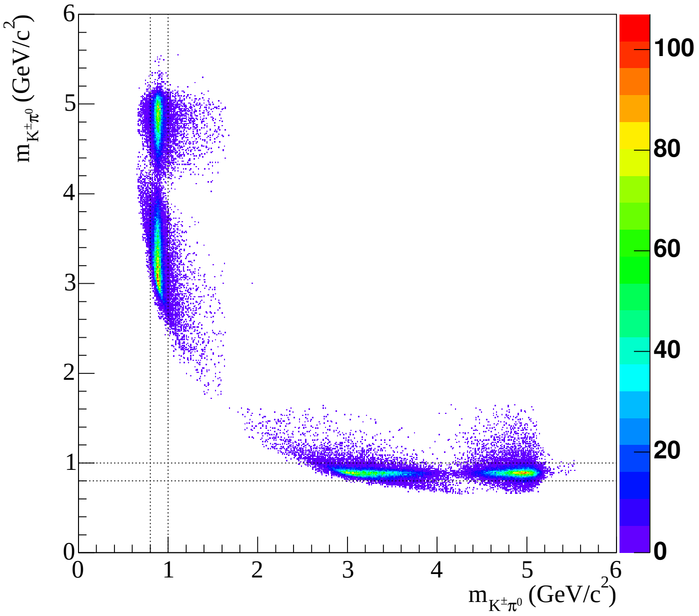
for any given candidate. For the case of the $\rho^{\pm}\pi^0$ analysis, the potential for overlap is much greater and an ambiguity is present depending on the values chosen for the mass range. However, the mass range used for the $\rho^{\pm}\pi^0$ analysis as described above precludes this ambiguity also. The issue of ambiguity is illustrated in Figure 3.8, where the square bounded by parallel dotted lines indicates the region whereby an ambiguity would be present.

The feature of no ambiguity in the resonance pair selection allows the Dalitz plot to be effectively 'folded' along the diagonal due to the final state being symmetric about this line.

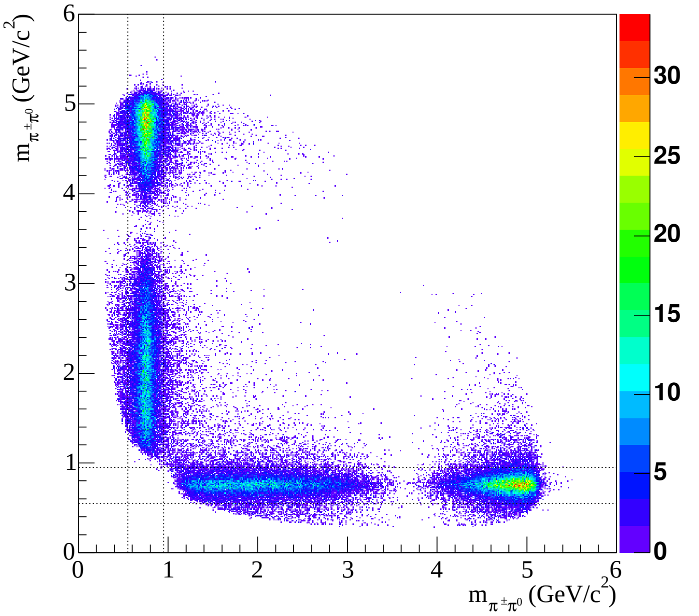
To summarise, the invariant mass cuts made for the $B^{\pm} \rightarrow K^{*\pm}\pi^0$ and $B^{\pm} \rightarrow \rho^{\pm}\pi^0$ analyses respectively are

- $0.8 < m(K^{\pm}\pi^0) < 1.0\,\mathrm{GeV}/c^2$
- $0.55 < m(\pi^{\pm}\pi^0) < 0.95\,\mathrm{GeV}/c^2$

In addition to being used in the selection, the resonance mass is used as a variable in the fit over the ranges given.



(a)



(b)

Figure 3.8: Distribution of invariant masses for the two $K^{\pm}\pi^0$ pairs (a) and $\pi^{\pm}\pi^0$ pairs (b) in the event. These plots are made for truth-matched signal $B^{\pm} \to K^{*\pm}\pi^0$ and $B^{\pm} \to \rho^{\pm}\pi^0$ respectively. In both cases the dashed lines representing the chosen mass cuts show that there is no appreciable ambiguity in choosing the lighter mass pair.

3.5 Continuum background fighting

3.5.1 Introduction

Events from light quark (u, d, s, c) continuum production represent the most significant type of background to this and most other charmless analyses. The construction of a suitable discriminating variable was considered as a means of identifying the signal B candidates efficiently, whilst rejecting events from the continuum. The problem of discrimination between two distinct classes of observations in multivariate data is well known in statistics and there exists a variety of different solutions. For the purposes of this analysis, two types of solution, or multivariate analyser (MVA), were considered:

i. Linear discriminant (LDA) in the form of either a Fisher [46] or Mahalanobis [47] discriminant.

ii. Nonlinear discriminant in the form of an artificial neural network (ANN).

In both cases the discriminant essentially maps a point in a multi-dimensional space to a point on a one-dimensional line, which is then treated as a single new variable. The mapping function which does this is derived with the use of *a priori* known samples (i.e. the MC simulated events) from the two classes of signal and background. These samples are referred to as training samples. The performance of the MVA is assessed with the use of two other independent samples of the same classes, which are referred to as testing samples.

The suitability of any potential discriminating variable was based on the criteria of simplicity and improvement in background rejection over an already existing Fisher discriminant, a modified CLEO Fisher variable, popular in other charmless 3-body analyses and described in further detail in [48]. This variable is a weighted linear combination of the summed momentum of the rest of the event seen in nine discrete cones centred around the thrust axis of the particles composing the B candidate, with additional inputs of the $\cos(\theta_{B,beam})$ (the cosine of the angle between the B direction and the beam axis) and $\cos(\theta_{T_B,beam})$ (the cosine of the angle between the thrust axis of the particles composing the B candidate and the beam axis). Although all comparisons of continuum rejection power were gauged with respect to the performance of the modified CLEO Fisher (MCF), this variable itself was allowed to feature as an input to the MVA.

The subject of statistical discrimination, classification and pattern recognition is too vast to be given justice here, and only a short summary of the two types of discriminant considered follows.

3.5.2 Linear discriminant

The principle behind a linear discriminant is to derive a linear function z of the n discriminating variables $x_1, ..., x_n$,

$$z = a_1 x_1 + ... + a_n x_n, \tag{3.9}$$

such that the ratio of the between-class variance of z to the within-class variance of z is maximised. The variable z is often referred to as Fisher's discriminant [46], or the first canonical variate. The vector $\mathbf{a} = [a_1, ..., a_n]$ can be regarded as defining the vector normal to a single hyper-plane dividing a multi-dimensional space into two separated regions. The optimal direction for $\mathbf{a}$ is that which maximises the separation of the projected class means while minimising the class variances. This is equivalent to finding $\mathbf{a}$ by maximising

$$V = \frac{\mathbf{a^T B a}}{\mathbf{a^T W a}}, \tag{3.10}$$

where $\mathbf{B}$ is the covariance matrix of the group means and $\mathbf{W}$ is the pooled within-groups covariance matrix:

$$\begin{aligned} \mathbf{B}_{ij} &= \sum_{classes,c} \frac{N_c}{N_{tot}} (\overline{x_i}^c - \overline{x_i}^{tot})(\overline{x_j}^c - \overline{x_j}^{tot}) & (3.11) \\ \mathbf{W}_{ij} &= \frac{1}{N_{tot}} \sum_c \sum_{x \in c} ({x_i}^c - \overline{x_i}^c)({x_j}^c - \overline{x_j}^c), & (3.12) \end{aligned}$$

where the indexes i, j run over the input variables, N_{tot} is the total number of observations of all classes and N_c is the number of observations in the training sample of each class. The vector $\mathbf{a}$ which maximises V (with value λ) of Eq.(3.10) is the principal eigenvector of

$$\mathbf{W}^{-1}\mathbf{B}\mathbf{a} = \lambda\mathbf{a} \tag{3.13}$$

which, in the case of only two classes, is given analytically by

$$\mathbf{a} = \mathbf{W}^{-1}(\overline{\mathbf{x_1}} - \overline{\mathbf{x_2}}). \tag{3.14}$$

There is a second form of linear discriminant, the *Mahalanobis* discriminant [47], which is also considered. It is similar to the Fisher, only it uses the inverse of the total covariance matrix $\mathbf{T} = \mathbf{B} + \mathbf{W}$ in place of $\mathbf{W}^{-1}$.

3.5.3 Artificial neural network

The second approach to the discrimination problem is to try a non-linear solution. The choice of discrimination function z used in this analysis defines the architecture of what is commonly termed a (single output) 'layered feed-forward artificial neural network' or 'multilayer perceptron' [49]:

$$z = \phi(\alpha_0 + \sum_j w_{j0}\phi(\alpha_j + \sum_i w_{ij}x_j)). \tag{3.15}$$

where $\phi(x)$ is called the *neuron activation function* and in this case is given by $[1+\tanh(x)]/2$, the index i runs over the input variables and j runs over the 'hidden nodes'. Figure 3.9 depicts a single-output, single hidden-layer feed-forward neural network with three inputs, four hidden nodes and one output. Layers in between the input and output layers are called 'hidden layers' since their outputs are not directly observable. The 'weights' w_{ij} and w_{j0} are determined from the training sample by minimising the error E:

$$E = \frac{1}{2N_p}\sum_{p=1}^{N_p}(z^{(p)} - t^{(p)})^2 \tag{3.16}$$

where t is the target value (in this case 1 for signal and 0 for background), z the artificial neural network output, N_p is the number of events in the training sample and p is the event or 'pattern' index.

There is no analytic method for minimising the error when the activation function is non-linear, instead numerical methods must be used. Minimising the error by adjusting the weights between the layers is called 'training' or 'learning'. The learning algorithms used here are of the gradient descent or *back-propagation* (BP) kind, which represent a simple but effective solution though not necessarily the fastest. The learning is done iteratively, with the vector of weights $\mathbf{w}(t+1)$ at the time step $(t+1)$ being adjusted by a quantity $\Delta\mathbf{w}(t)$, given by

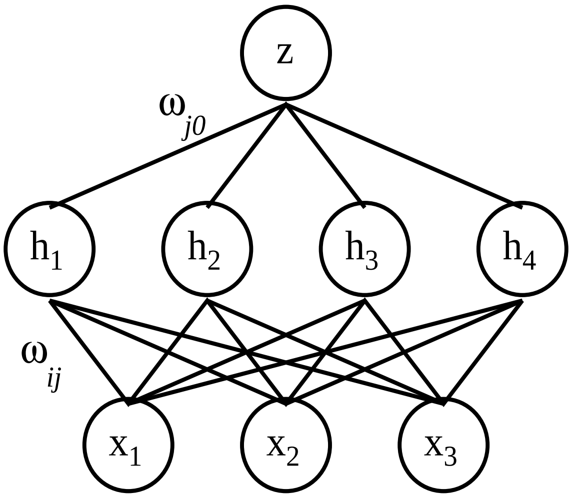


Figure 3.9: Architecture of a single-output, single hidden-layer feed-forward neural network.

$$\Delta\mathbf{w}(t) = -\lambda\nabla_w E(t), \tag{3.17}$$

where λ is a constant quantity called the learning rate, which can be tuned to the specific training sample depending on the scale of the inputs. The quantity ∇_w represents the vector of derivatives $(\partial/\partial w_{ij})$. If λ is too big the ANN may never settle, but could oscillate back and forth across the minima. To avoid this a 'momentum' term is added which is fractionally $(\alpha < 1)$ proportional to the previous change,

$$\Delta\mathbf{w}(t) = -\lambda\nabla_w E(t) + \alpha\Delta\mathbf{w}(t-1), \tag{3.18}$$

thus if the previous change was large, so too will the next one be.

In practice the error is not calculated for the entire training sample, but on a subset of data and the updating is done at the end of this error calculation; the period between updating is called an epoch. To help avoid settling on local minima or to move out of a region of flatness in the error gradient, a Gaussian noise term $\mathbf{\Gamma}$ can be added to the weight vector updating term after each epoch

$$\Delta\mathbf{w}(t) = -\lambda\nabla_w E(t) + \alpha\Delta\mathbf{w}(t-1) + \mathbf{\Gamma}. \tag{3.19}$$

This is the principle behind *Langevin* updating [50]. Another learning algorithm investigated is the *Manhattan* [50] updating given by the rule

$$\Delta\mathbf{w}(t) = -\lambda.\text{sign}[\nabla_w E(t)]. \tag{3.20}$$

We move now to discuss the implementation of these discriminant techniques to this analysis.

3.5.4 Selecting the best input variables

As has been discussed earlier, continuum events are betrayed by their two-jet-like event topology. This ultimately holds the key to the discriminating power of any variable constructed. Variables initially chosen as potential inputs to the MVA are therefore event-shape variables. To determine which inputs provide the greatest discrimination between signal and continuum, a statistic called the discrimination power is compared. It is defined by the quantity

$$\text{Discrimination power (variable } i) = \frac{\mathbf{B}_{ii}}{\mathbf{W}_{ii}} \tag{3.21}$$

where $\mathbf{B}$ and $\mathbf{W}$ are the between- and within- class covariance matrices respectively of Eqs.(3.11) and (3.12) (there are only two classes, signal and background). This is essentially a comparison between the averaged spread within a subgroup to the spread over the whole population. This quantity is cross-checked against a similar quantity termed the signal discrimination power, defined as

$$\text{Signal discrimination power (variable } i) = \frac{(\overline{x_i}^{sig} - \overline{x_i}^{bkg})^2}{var(x_i{}^{sig})} \tag{3.22}$$

which is a measure of the relative distance between the means. The higher the value, the more useful the quantity is for discrimination.

Several variables were considered for inclusion in the initial investigation. There are variables that are either calculated using tracks and neutral objects associated with the B candidate (variables with a (B) suffix) or are from tracks and neutrals not associated with the B, so coming from the ROE. The full list of variables considered is given below:

- Aplanarity (ROE), aplR.
- Sphericity (ROE), sphR.
- Sphericity (B), sphB.
- Thrust(B), thrustB.
- Fox-Wolfram moments (ROE) fw2R, the ratio of the 2^{nd} moment to the 0^{th}, and fw4R, the ratio of the 4^{th} to the 0^{th}.
- fw2, the ratio of the 2^{nd} to the 0^{th} Fox-Wolfram moments for the whole event.
- Monomials L_0, L_2 (see below).
- Modified CLEO Fisher, 9 CLEO cones $+\cos(\theta_{B,beam}) + \cos(\theta_{T_B,beam})$.
- cos(SN), cosine of the angle between sphericity axis of the ROE and normal to B decay plane.
- cos(TN), cosine of the angle between thrust axis of the ROE and normal to B decay plane.
- cos(SB), cosine of the angle between sphericity axis of the ROE and B direction.
- cos(TB), cosine of the angle between thrust axis of the ROE and B direction.

- cos(TTB), cosine of the angle between thrust axis of the ROE and the B thrust axis.
- TFlv, the output of the Moriond Tagger [51].

The monomials L_0 and L_2 are variables similar to the Legendre polynomials and are defined as:

$$L_0 = \sum_i^{ROE} p_i, \qquad L_2 = \sum_i^{ROE} p_i \times \frac{1}{2}(3\cos^2(\theta_i) - 1) \tag{3.23}$$

where p_i and θ_i are the momentum magnitude and polar angle of tracks and clusters summed over candidates in the rest of the event.

Tables 3.4 and 3.5 show the values for the discrimination power and signal discrimination power respectively, for the two signal modes $B^\pm \to \rho^\pm\pi^0$ and $B^\pm \to K^{*\pm}\pi^0$, and the two non-resonant modes $B^\pm \to \pi^\pm\pi^0\pi^0$ and $B^\pm \to K^\pm\pi^0\pi^0$. All these modes are compared individually to the same sample of MC simulated continuum events. Each of the B samples is an equal-sized collection of truth-matched signal events, which have passed the same full preselection cuts and best candidate selection (described below in Section 3.6). Variables which have discrimination values different from zero are considered to be potentially useful for inclusion. The uncertainty in the values is at the level of 10^{03}.

Both methods of highlighting the variables with the greatest discrimination power give the same results, with seven highlighted variables standing out from the rest. They are L_0, L_2, MCF, cos(TTB), thrustB, sphB and fw2. It was also decided to use TFlv although it would appear to have a low discrimination power. It should be emphasised that the statistic used to evaluate the discrimination power assumes Gaussian distributions for its optimality. The shape of TFlv is such that the statistic does not do it justice and the neural net discrimination surface can probe its distribution in a way a linear discriminant cannot.

Before any further investigations proceeded it was desired to have input distributions which were as similar as possible for both the signal modes and the *non-resonant*

Variable	$\pi^{\pm}\pi^0\pi^0$	$\rho^{\pm}\pi^0$	$K\pi^0\pi^0$	$K^*\pi^0$
sphR	0.00	0.00	0.00	0.00
sphB	**0.26**	**0.13**	**0.32**	**0.14**
thrustB	**0.33**	**0.01**	**0.39**	**0.04**
L0	**0.06**	**0.06**	**0.04**	**0.07**
L2	**0.57**	**0.54**	**0.50**	**0.44**
Fisher	**0.60**	**0.52**	**0.55**	**0.49**
cosSN	0.00	0.00	0.00	0.00
cosTN	0.00	0.00	0.00	0.00
cosTB	0.00	0.00	0.00	0.00
cosSB	0.00	0.00	0.00	0.00
cosTTB	**0.04**	**0.05**	**0.04**	**0.04**
aplR	0.00	0.00	0.00	0.00
fw2R	0.01	0.01	0.00	0.00
fw4R	0.00	0.00	0.01	0.00
fw2	**0.10**	**0.17**	**0.10**	**0.12**
TFlv	**0.01**	**0.00**	**0.00**	**0.00**

Table 3.4: Values of the variables' discrimination power for truth-matched signal and non-resonant modes.

	$\pi\pi^0\pi^0$	$\rho^{\pm}\pi^0$	$K\pi^0\pi^0$	$K^*\pi^0$
sphR	0.01	0.02	0.00	0.00
sphB	**0.56**	**2.96**	**0.70**	**2.81**
thrustB	**0.79**	**0.06**	**0.90**	**0.25**
L0	**0.26**	**0.28**	**0.16**	**0.30**
L2	**2.58**	**2.49**	**2.06**	**1.83**
Fisher	**2.86**	**2.55**	**2.50**	**2.29**
cosSN	0.00	0.00	0.00	0.00
cosTN	0.00	0.00	0.00	0.00
cosTB	0.02	0.00	0.00	0.00
cosSB	0.00	0.00	0.01	0.00
cosTTB	**0.21**	**0.23**	**0.22**	**0.22**
aplR	0.00	0.00	0.01	0.01
fw2R	0.04	0.05	0.01	0.01
fw4R	0.00	0.00	0.02	0.01
fw2	**0.50**	**0.77**	**0.42**	**0.56**
TFlv	**0.02**	**0.00**	**0.00**	**0.00**

Table 3.5: Values of the variables' signal discrimination power for the truth-matched signal and non-resonant modes.

decay channels $B^{\pm} \to K^{\pm}\pi^0\pi^0$ and $B^{\pm} \to \pi^{\pm}\pi^0\pi^0$. The level of non-resonant events in the final sample is unknown, and must be estimated. As will be discussed in the next chapter, the numbers of non-resonant decays expected in the final data samples are estimated from the data using sidebands in the Dalitz plane and the method of estimation uses the continuum discrimination variable in the same way as the resonant signal analyses. If for example the neural network was optimised for the resonant channels exclusively, and was non-optimal for the non-resonant decays, then extrapolation of non-resonant events from sidebands into the signal regions becomes more involved and will inevitably introduce additional systematic uncertainties. It was desired therefore that the distribution of the MVA for signal be insensitive to the position of the decay in the Dalitz plane. Variables involving resonance-specific dependencies (such as $m(K^{*\pm}, \rho^{\pm})$, $\cos(\theta_H)^2$ etc.) were excluded from consideration for this reason. While most of the selected variables use the tracks and clusters from the rest of the event, which should be the same independent of signal final state, sphB and thrustB are built around the B decay products which vary significantly between the resonant and non-resonant modes.

Figure 3.10 shows the different distributions for the signal modes for each of the input variables. While the majority of the variables are in very good agreement, sphB and thrustB show significant differences due to the resonant and non-resonant nature of the decays. Consequently these variables are removed from further consideration. The final choice of variables is narrowed to six: L_0, L_2, MCF, cos(TTB), fw2 and TFlv. It should be noted that although the variables are shown for the $B^{\pm} \to K^{*\pm}\pi^0$ modes, the $B^{\pm} \to \rho^{\pm}\pi^0$ are very similar and the plots can be considered representative of both samples.

3.5.5 Correlation of variables with ΔE, m_{ES}.

As the output of any discriminant will be used in a maximum likelihood fit along with m_{ES} and ΔE, it is important that any input variables to the discriminant are not correlated with m_{ES} and ΔE otherwise this could result in the discriminant output also being correlated. The likelihood fit construction, described in Chapter

[2] $\cos(\theta_H)$ is defined as the cosine of the angle between the momentum of the track and the momentum of the B as measured in the rest frame of the resonance.

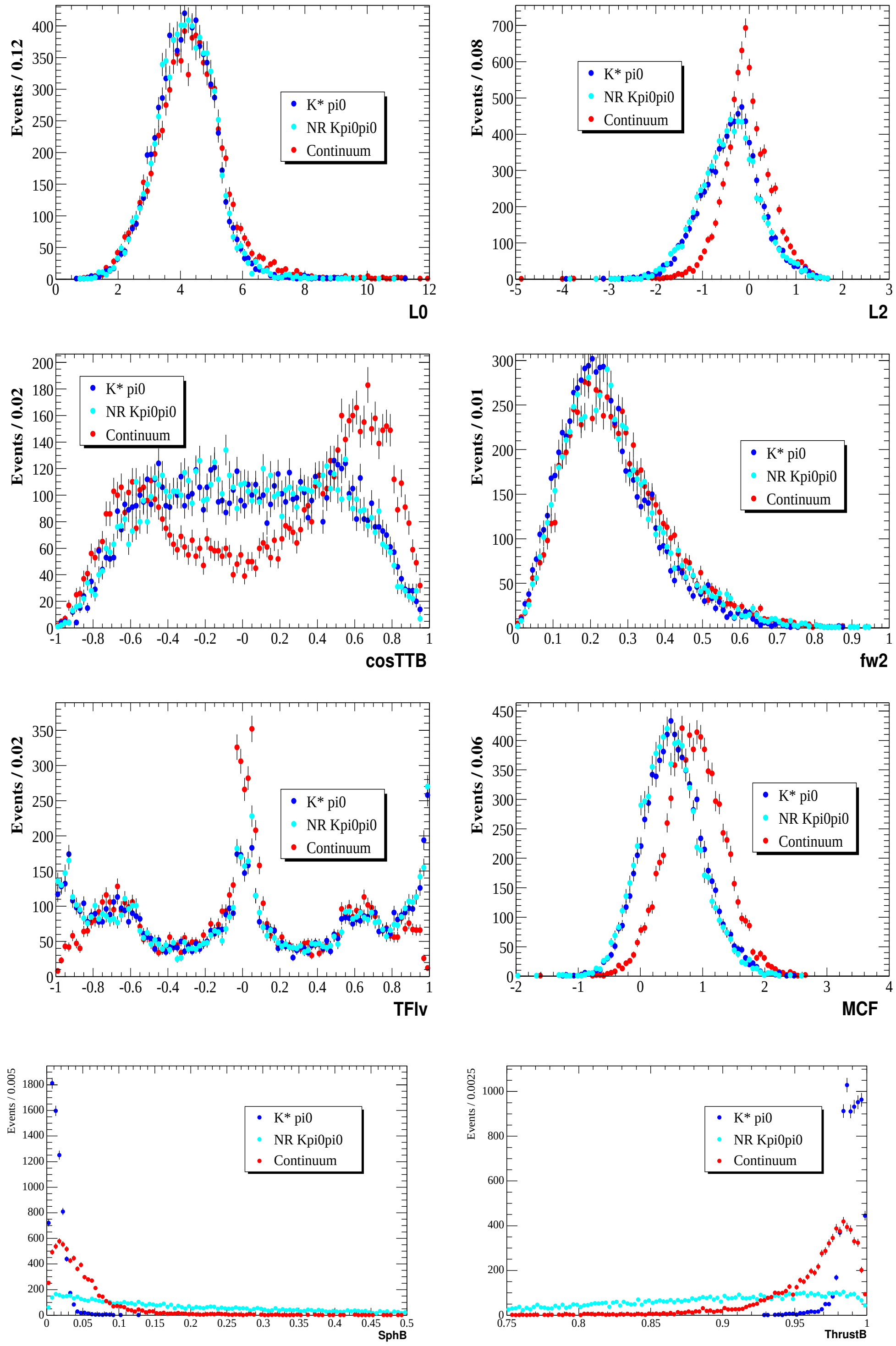


Figure 3.10: The distributions of the discriminating variables for the truth-matched $K^{*\pm}$, non-resonant $K^{\pm}\pi^0\pi^0$ and continuum. The distributions (reading from top left) are L_0, L_2, cos(TTB), fw2, TFlv, MCF, sphB and thrustB. The distributions are similar for both resonant and non-resonant modes, except for the last two variables which is the reason for their subsequent exclusion.

4, assumes orthogonality of the probability density functions (PDFs) describing the distribution of the input variables. In the cases where the input variables are correlated, the PDFs must be constructed to account for the correlation. In using input variables that uncorrelated, one can describe the total PDF over all input variables as a product of orthogonal one-dimensional PDFs over the individual inputs which thereby reduces the complexity of the fit construction.

The correlations of these variables with m_{ES} and ΔE are shown in Tables 3.6 and 3.7. The statistic used was Pearson's product moment correlation coefficient R, given by

$$R = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}. \tag{3.24}$$

In all cases we find that there are no correlations at a level that might cause concern.

	m_{ES}	ΔE
L_0	-0.001	-0.016
L_2	-0.014	0.020
MCF	0.028	-0.004
cosTTB	0.022	0.002
TFlv	-0.025	-0.001
fw2	-0.020	0.031

Table 3.6: Values of the variable correlations with ΔE, m_{ES} for the truth-matched $K^{\pm}\pi^0\pi^0$.

	m_{ES}	ΔE
L_0	-0.005	-0.007
L_2	-0.018	-0.025
MCF	0.021	-0.030
cosTTB	0.009	-0.038
TFlv	-0.035	0.021
fw2	0.020	0.057

Table 3.7: Values of the variable correlations with ΔE, m_{ES} for the truth-matched $K^{*\pm}\pi^0$.

3.5.6 MVA designs

For both the linear discriminant types, custom software was written to interface with ROOT.

The *ANN* software used is based on JetNet v.3.5 [52] with a front-end interface to ROOT [53]. The neural network architecture is a three layer (1 input, 1 hidden, 1 output) perceptron with 1 output node. The number of hidden nodes used was varied from 4 to 10 and was found to not have any significant impact on the discrimination performance for any one particular set of inputs. The number of possible inputs is a maximum of 6, so in principle a 7-node hidden layer is the minimum node requirement to completely close off a volume in the input space[3]. From the shape of the input distributions, it is apparent that the minimum requirement to completely close off a volume is unnecessary, thus in keeping with the adopted principle of simplicity in design a six-node hidden-layer was chosen for the final configuration.

Two types of training algorithm were investigated, namely standard back-propagation of the error with either Langevin or Manhattan updating.

3.5.7 MVA sample composition

The samples used for training the *ANN* and for constructing the LDA were identical and were composed as follows:

- **Signal** : Truth matched signal B candidates from each of the four categories $B^{\pm} \to \rho^{\pm}\pi^{0}$, $B^{\pm} \to \pi^{\pm}\pi^{0}\pi^{0}$, $B^{\pm} \to K^{*\pm}\pi^{0}$ and $B^{\pm} \to K^{\pm}\pi^{0}\pi^{0}$ were chosen at random from the MC samples and concatenated to a list in a random order. As the training of the *ANN* updates the weight vectors after every epoch (typically 10 events in this investigation), the updated weight vector should not be biased towards any particular signal mode so training on all modes has to be performed simultaneously. Each mode is also weighted equally by randomising the selection process. The sample size is 14000, with the sample split evenly in two groups of 7000 events, one for training and one for testing.

[3]By examination of Eq.(3.15), one can regard a single hidden node in the *first layer* as being a function related to the orientation of a single hyper-plane in the *input space*. Thus to define a closed volume in the N dimensional input space, one needs a minimum of N+1 hyper-plane surfaces (of dimension N-1).

- **Background** : As the cross sections for continuum event production are known [20], the proper mixture of MC simulated u, d, s, c and τ events is respected in the composition of the background training sample. The sample size for background, like signal, is 7000 training events, 7000 testing events.

3.5.8 Optimal ANN configuration

Figure 3.11 shows the performance of 10 of these combinations of input variables for a Manhattan trained variable. The binary code to identify the efficiency curves is indicative of inclusion (1), or exclusion (0) for the variables in the respective order: MCF, L_0, L_2, cos(TTB), TFlv, and fw2. For example, the pattern 011010 means L_0, L_2 and TFlv were all included while all others were left out. Out of the six variables remaining, it was found that inclusion of the CLEO Fisher or the pair (L_0, L_2) or both was consistently found to have the most powerful discrimination. As a result all the combinations shown feature one of these inputs.

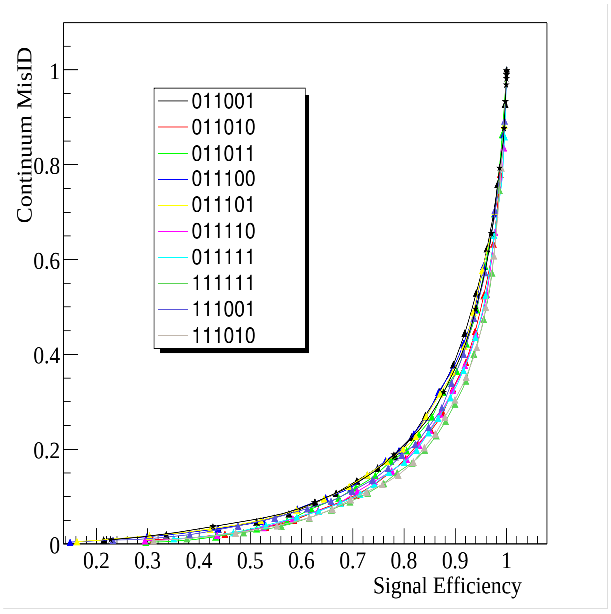


Figure 3.11: Performance plot of various *ANN* configurations showing the continuum selection efficiency (y axis) against signal selection efficiency (x axis) for cuts applied to the output from the *ANN*. The binary code is explained in the text. The *ANN* in its final form corresponds to the input combination 011010.

It was found that the *ANN* does best when L_0, L_2 and MCF are all used in combination and that the inclusion of TFlv always yields better performance than when it is excluded for any given combination. The most powerful combination is actually the inclusion of all the variables, but it is only very marginally better than the combination of MCF, L_0, L_2 and TFlv. However, after further investigations uncomfortably high correlations with m_{ES} were seen with an *ANN* which included the MCF as an input, particularly in the tail of m_{ES}. This correlation was not seen so strongly in the global correlations between the individual input variables. From the design objective of simplicity, the final input configuration is therefore a combination of the three variables $\mathbf{L_0}, \mathbf{L_2}$ and **TFlv**. The training algorithm chosen for the *ANN* is the Manhattan updating. The output of the optimal *ANN* for signal and continuum events is shown in Figure 3.12. As one can see, signal events give a response that is sharply peaked near 1, while background continuum events give a distribution which is sharply peaked near 0.

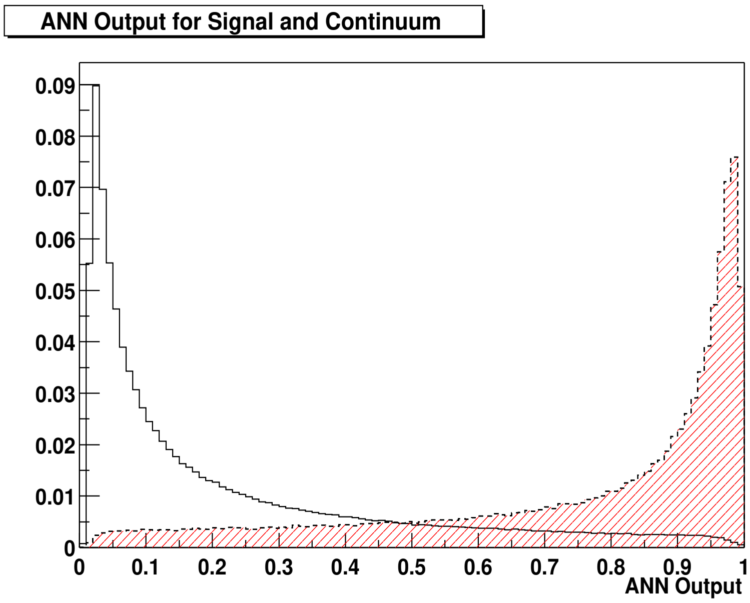


Figure 3.12: The *ANN* output for signal $K^{*\pm}\pi^0$ (hatched/red histogram) events and for continuum events (clear histogram).

3.5.9 Comparison of *ANN* and LDA

The performance of the selected *ANN* combination was compared to a linear discriminant to see which would provide the best discrimination. The output of the the linear discriminant was in two forms, the first being a new Fisher type discriminant and the second one based on the Mahalanobis technique. The LDA output for each variable for signal and background is shown in Figure 3.13.

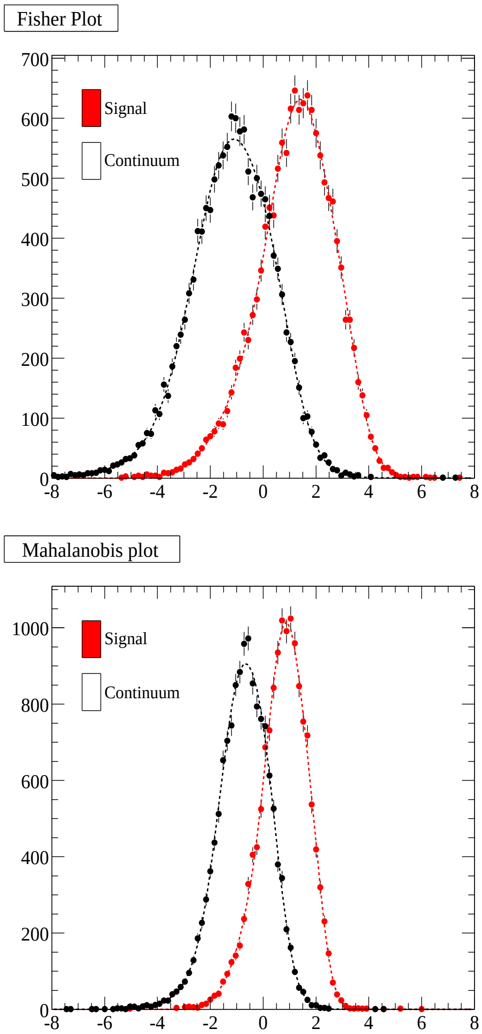


Figure 3.13: Examples of the output for signal and background for the two linear discriminants. The left hand plot shows the new Fisher discriminant and the right hand plot shows the Mahalanobis based discriminant. Both discriminants are using the combination of inputs which was found to be most effective for the neural network (011010).

In all cases it was found that the linear discriminant variables performed worse than the *ANN* with the equivalent input combination. As a demonstration, the Mahalanobis and new Fisher solutions for the final four input combinations along with their comparison with the *ANN* are shown in Figure 3.14. It was therefore decided that the *ANN* as described above would be used as the variable for continuum rejection in this analysis.

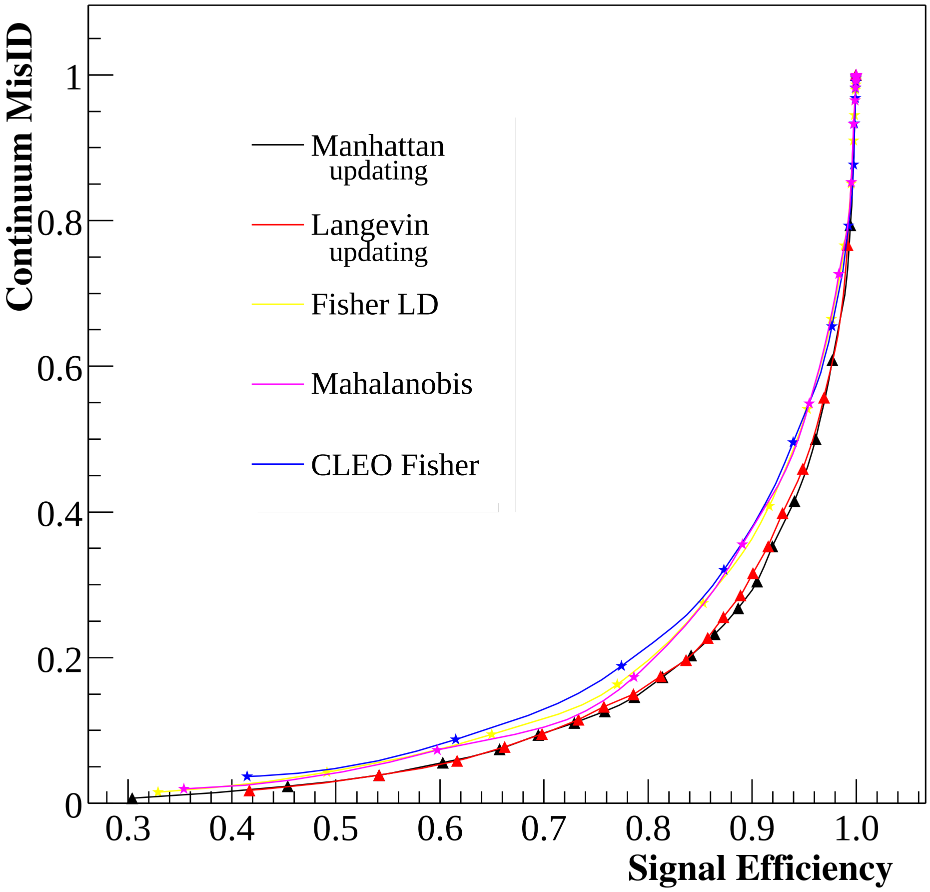


Figure 3.14: Comparison of the discriminating power of both linear discriminant variables, the *ANN* variable trained with the Langevin and Manhattan algorithms and the original CLEO based Fisher. The plot is for a combined signal sample of $B^{\pm} \to K^{\pm}\pi^0\pi^0$, $B^{\pm} \to \pi^{\pm}\pi^0\pi^0$, $B^{\pm} \to \rho^{\pm}\pi^0$ and $B^{\pm} \to K^{*\pm}\pi^0$ and a background sample of MC simulated continuum events. Each plot shows the continuum selection efficiency (y axis) against signal selection efficiency (x axis) for cuts applied to the output discriminating variable. The closer the curve is to the bottom right corner, the more powerful is the discrimination.

3.5.10 MVA conclusions

In this section we have investigated two different forms of discriminant to separate signal events from the dominant continuum background. One is a neural network based discriminant and the other is a linear discriminant. Both have used event shape variables as inputs, which have been carefully selected to give maximum discriminating power. We find that the *ANN* consistently gives the better performance and is selected as the background fighting variable.

The *ANN* is now used in two ways within the analysis. We first apply a cut at *ANN*

> 0.25 ($B^{\pm} \to K^{*\pm}\pi^0$) and $ANN > 0.4$ ($B^{\pm} \to \rho^{\pm}\pi^0$) in the neural network output (see Figure 3.12) to deselect the majority of all continuum events while retaining good signal efficiency. These cuts are determined from an optimisation study that is fully explained in Chapter 4, as its explanation requires the introduction of the maximum likelihood fit. Secondly the variable is transformed before being used as an input to the fit. The reasons for this transformation are outlined below.

Transformation of *ANN*

The distribution of the neural network output for true B events is sharply spiked around 1. This raises a practical issue in the modelling, specifically in how to describe its shape adequately to form an accurate probability density function for a given B mode (especially when MC statistics for some modes are small).

To try to avoid this issue, the ANN output was transformed using the following *ad hoc* function:

$$\text{Transformed } ANN \;=\; ANN_{Tran} \;=\; 1 - \left(\frac{1}{\pi} \times \arccos(ANN)\right) \tag{3.25}$$

This leads to a much smoother distribution, which also has an easy and well-modelled parametric description. The details of the parametric description for the PDFs are left until Chapter 4.

3.6 Multiple candidates

For events which pass the preselection, it is usual to have more than one B candidate per event. The average multiplicity for the signal distributions is around 1.5 candidates per event. This combinatorial self-cross-feed (SCF) background can be problematic. If one chooses to be inclusive of this background and retain as much information as possible, one must model it in the fit, taking into account the multiplicity distributions for *all* modes. However this modelling is entirely reliant on the MC simulation, is difficult to verify on data and adds a level of complexity to the fit that is unlikely to be helpful in measuring what we are essentially after: the number of signal events.

It is thus entirely reasonable to remove this appreciable combinatorial background by selecting only one 'best' candidate per event. However, one must be careful to avoid potential bias, and thus any choice must be relatively independent of the variables used in the fit. Multiple candidates arise in the signal distributions broadly by two possible ways:

- Swapping out a soft π^0 from the resonance (the π^0 which comes directly from the B decay is always reasonably hard and rarely misidentified or swapped out) with one from the ROE. This affects both the $B^{\pm} \to \rho^{\pm}\pi^0$ and $B^{\pm} \to K^{*\pm}\pi^0$ analyses.
- Swapping out the track from the resonance with one from the ROE, usually when it is slow. This only really affects the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis - the $B^{\pm} \to K^{*\pm}\pi^0$ puts PID requirements on the track to be a kaon, for which there are a lot fewer in the event to swap out and hence less chance to reconstruct a good B. Also the PID requirements effectively put a lower limit on the momentum allowed for the track, thereby rejecting many of the soft tracks from the event whether they are true kaons or not.

Motivated by this, the following strategies for choosing the best candidate were investigated:

- A B candidate χ^2 formed from the π^0 masses.
- A B candidate likelihood formed from the π^0 masses (the distinction between these two is made below).
- A choice based on the cosine of the helicity angle, $\cos\theta_{Hel}$ of the ρ resonance.
- A random candidate.

The choice of $\cos\theta_{Hel}$ is only applied to the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis. This is because the only appreciable source of multiple candidates in the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis is from the fake π^0 candidates or swapping a genuine slow π^0 from the ROE. In this case $\cos\theta_{Hel}$ really does not help discriminate against the SCF component. It is only when one has additional multiplicities associated with the track that the helicity can discriminate effectively between mistag events. This is discussed below.

Candidate χ^2

The simplest form of discrimination against fake π^0 candidates is to compare the reconstructed π^0 candidates to the nominal π^0 mass. A B candidate χ^2 is formed:

$$\chi^2_{Cand.} = \frac{(M_{\pi^0_1} - M^{PDG}_{\pi^0})^2}{\sigma^2(M_{\pi^0})} + \frac{(M_{\pi^0_2} - M^{PDG}_{\pi^0})^2}{\sigma^2(M_{\pi^0})} \tag{3.26}$$

where $\sigma^2(M_{\pi^0})$ is the nominal π^0 resolution of the *BABAR* EMC, and is a constant quantity. The candidate chosen is the candidate with the lowest value of χ^2_{Cand}. This amounts to effectively taking the χ^2_{Cand} of the π^0 from the resonance as most multiple candidates share the hard π^0 from the B. Selecting on the π^0 quality will not be very effective if the backgrounds are mainly real π^0 mesons.

Candidate likelihood

An evolution of the above algorithm is to make a likelihood for the π^0 mass, rather than just selecting the candidate with mass closest to the nominal value. This effectively means evaluating the constant resolution $\sigma^2(M_{\pi^0})$ in the Eq.(3.26) above as a function of the π^0 momentum. This is obtained by fitting an appropriate function to the $\gamma\gamma$ invariant mass for a sample of truth-matched π^0 candidates, in bins of momentum. The functional form used is the 'Novosibirsk' function [54], which is essentially a Gaussian with a tail. The likelihood $\mathcal{L}_{Cand.}$ is formed as

$$\mathcal{L}_{Cand.} = \mathcal{L}^{\pi^0_1}(M(\mathrm{p}_1), \sigma(\mathrm{p}_1)) \cdot \mathcal{L}^{\pi^0_2}(M(\mathrm{p}_2), \sigma(\mathrm{p}_2)).$$

The candidate chosen is that with the highest likelihood.

Candidate ρ - $\cos\theta_{Hel}$

In contrast, for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis, a selection based on the π^0 mass is of no use when the multiplicity is due to the track. In such cases it is possible to reconstruct the B candidate with the opposite charge from the true decay. It is particularly important to reduce this mistag fraction as it has the effect of diluting the asymmetry measurement. A selection algorithm based on the helicity was identified as having potential to reject this mistag SCF component, for the mistag events are all clustered at high positive values of $\cos\theta_{Hel}$, as can be seen in Figure 3.15.

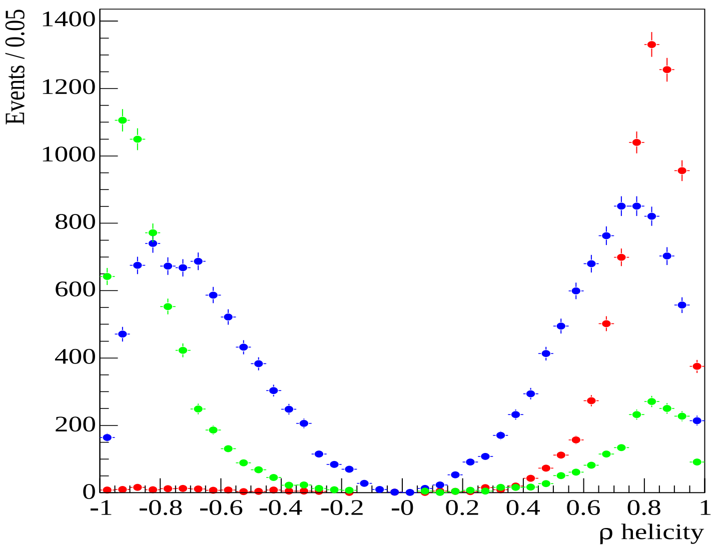


Figure 3.15: Plot of the ρ $\cos\theta_{Hel}$ for signal $\rho^{\pm}\pi^{0}$ (blue), correct-charge SCF (green) and wrong-charge SCF (red). For correctly reconstructed signal the distribution should be parabolic in $\cos\theta_{Hel}$. The fall-off in this functional form at the edges is an efficiency effect.

The algorithm to select between two ρ candidates with values of $\cos\theta_{Hel}$ as h_1 and h_2 is

- If $h_1 > 0$ and $h_2 > 0$ select the candidate with the smaller value.
- If $h_1 > 0$ and $h_2 < 0$ select the candidate with the negative value.
- If $h_1 < 0$ and $h_2 < 0$ select the candidate with the least negative value.

Random candidate

These algorithms are to be compared to choosing a candidate randomly. If the average multiplicity is m for events which have a truth match, and one does worse than $1/m$ on average with the adopted algorithm, it would clearly be better to choose the candidate randomly.

3.6.1 Final algorithm

The algorithm chosen was based on the ability to retain as many truth-matched candidates as possible in the case of the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, and reject as many mistag candidates as possible in the case of the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis. A summary of the effective performance of the strategies is shown in Table 3.8.

Summary of Best Candidate Strategies for $B^{\pm} \to K^{*\pm}\pi^0$			
Algorithm	% of events which have TM before selection	% of events (candidates) which have TM after selection	% of mistag candidates
$\chi^2_{Cand.}$	68.3	64.4	$< 1\%$
$\mathcal{L}_{Cand.}$	68.3	64.6	$< 1\%$
random	68.3	60.7	$< 1\%$
Summary of Best Candidate Strategies for $B^{\pm} \to \rho^{\pm}\pi^0$			
$\chi^2_{Cand.}$	62.5	54.9	7.6
$\mathcal{L}_{Cand.}$	62.5	55.0	7.7
$\rho \cos\theta_{Hel}$	62.5	55.3	7.2
random	62.5	51.1	7.7

Table 3.8: Table summarising the performance of the various best candidate algorithms. In all cases the algorithms did better than a random selection. TM = truth-match.

For the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, the algorithm chosen was the $\chi^2_{Cand.}$. Although the $\mathcal{L}_{Cand.}$ performed slightly better, the difference was marginal enough to raise the issue of uncertainties in the data:MC simulation differences for the π^0 mass resolution. The likelihood has a preference to select low energy π^0 candidates where the resolution is superior but also where the data:MC simulation discrepancies are larger. It was decided that the most transparent and robust strategy was to use $\chi^2_{Cand.}$.

For the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis the $\cos\theta_{Hel}$-based selection was the preferred strategy for the reason that it was the best at reducing the fraction of mistag candidates being selected.

In both studies the random selection performs better than the naive estimate of $1/m$, where m is the multiplicity (see Table 3.9). This is because most events have only one candidate per event, and the multiplicity is largely due to a small number of events having a large number of candidates. Thus although the gains from the best candidate selection may appear modest, it is in these rarer instances of high multiplicity that they have the advantage over the random selection.

This investigation was carried out with a selection cut-off in the ANN variable of 0.2. As will be described in the next chapter, this cut-off is further optimised for each analysis. For the $B^{\pm} \to K^{*\pm}\pi^0$ analysis this cut-off is $ANN > 0.25$ and for the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis it is $ANN > 0.40$. This has a small effect on the numbers that appear in Table 3.8. After all selections, the truth-match, correct-charge SCF fraction and mistag fractions for the two analyses are:

$B^{\pm} \to K^{*\pm}\pi^0$

- Truth-match fraction: 64.5%
- Correct charge SCF fraction: 35.5%
- Mistag fraction: negligible

$B^{\pm} \to \rho^{\pm}\pi^0$

- Truth-match fraction: 56.04%
- Correct charge SCF fraction: 36.93%
- Mistag fraction: 7.03%

One final word must be said about the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis. The construction of the event likelihood (Section 4.3.1) is done in such a way as to consider the number of signal events reconstructed with the wrong charge as being *proportional* to the number of signal events reconstructed with the correct charge. This requires a further set of numbers to be defined, the fraction of *correct charge* signal events that are SCF, f_{SCF}, and the *ratio of wrong charge to correct charge signal events*, f_{Mistag}. These are simply obtained from the numbers already given:

$$\begin{aligned} f_{SCF} &= 36.9/(36.9+56.0) &= 39.7\% \\ f_{Mistag} &= 7.03/(36.9+56.0) &= 7.6\% \end{aligned}$$

3.7 Summary of cuts and efficiencies

The following tables show a summary of the cuts and the efficiencies for signal, continuum and generic $b \to c$ MC simulation and off-resonance and on-resonance data after all cuts for the two analyses. Multiplicity here means the number of B candidates per event.

Signal efficiencies (%)		
Cut	$B^{\pm} \to K^{*\pm}\pi^0$	$B^{\pm} \to \rho^{\pm}\pi^0$
`CPi0Pi0` Tag (Multiplicity)	47.2 (5.19)	50.9 (5.30)
$\cos\theta^{B}_{\mathrm{Sph}} < 0.9$	87.8	87.1
$\|\Delta E\| < 0.25$ $5.19 < m_{ES} < 5.3$	94.6	91.1
$\cos\theta^{\gamma}_{Hel}$	88.5	86.6
$0.01 < LAT_{\gamma} < 0.6$	94.0	94.7
$0.11 < M_{\pi^0} < 0.16$	96.6	96.6
$0.35 < \theta_{\gamma} < 2.39$	96.5	96.5
Electron Veto	99.8	99.4
PID	86.7	87.0
$0.8 < m_{K^*} < 1.0\,\mathrm{GeV}/c^2$	73.9	
$0.55 < m_{\rho} < 0.95\,\mathrm{GeV}/c^2$		76.7
$ANN > 0.2$	93.9	93.6
$-0.2 < \Delta E < 0.2$ $5.2 < m_{ES} < 5.3$ $ANN > 0.25$ (0.4)	91.7	80.2
Total (Multiplicity)	16.48 ± 0.09 (1.51)	15.41 ± 0.09 (1.75)
Truth Match	64.50 ± 0.20	56.40 ± 0.60

Table 3.9: The efficiencies of the selection cuts for the signal modes, as obtained from the MC simulation. The cuts are performed hierarchically (in the order given in the table), with the values being the efficiency of the cut with respect to the remaining sample of events.

Data efficiencies (%)		
Cut	Off–Resonance	On–Resonance
`CPi0Pi0` Tag (Multiplicity)	– (2.32)	– (2.29)
$\cos\theta^{B}_{\mathrm{Sph}} < 0.9$	55.5	55.8
$\lvert\Delta E\rvert < 0.25\,\mathrm{GeV}$ $5.19 < m_{ES} < 5.3\,\mathrm{GeV}/c^2$	67.6 [54.9]	67.2 [54.5]
$\cos\theta^{\gamma}_{Hel}$	86.8 [87.8]	86.2 [87.8]
$0.01 < LAT_{\gamma} < 0.6$	85.1 [85.4]	83.4 [85.7]
$0.11 < M_{\pi^0} < 0.16\,\mathrm{GeV}/c^2$	91.9 [92.0]	91.8 [92.2]
$0.35 < \theta_{\gamma} < 2.39$	93.7 [93.8]	93.9 [93.7]
Electron Veto	98.6 [98.7]	98.5 [98.6]
PID	23.9 [58.7]	24.3 [58.5]
$0.8 < m_{K^*} < 1.0\,\mathrm{GeV}/c^2$ $0.55 < m_{\rho} < 0.95\,\mathrm{GeV}/c^2$	17.8 [32.0]	17.5 [31.9]
$ANN > 0.2$	46.9 [43.7]	48.8 [44.2]
$-0.2 < \Delta E < 0.2\,\mathrm{GeV}$ $5.2 < m_{ES} < 5.3\,\mathrm{GeV}/c^2$ $ANN > 0.25$ [> 0.4]	60.0 [38.3]	60.9 [39.2]
Total Events (Multiplicity)	1731 [3698] (1.28)	23465 [47608] (1.29)

Table 3.10: The efficiencies of the selection cuts for the data samples. As the ΔE selection assumes a different mass hypothesis for the track in each analysis, the analyses begin to diverge and the efficiencies for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis are indicated in brackets and coloured red for clarity. The cuts are performed hierarchically (in the order given in the table), with the values being the efficiency of the cut with respect to the remaining sample of events. The final numbers of events passed is given at the bottom.

Light quark continuum and generic $b \to c$ efficiencies (%) (MC simulation)			
Cut	generic $b \to c$ ($B^{\pm}$)	generic $b \to c$ (B^0)	uds
`CPi0Pi0` Tag (Multiplicity)	$(6.27) \times 10^{-2}$ (2.27)	$(3.23) \times 10^{-2}$ (2.19)	1.52 (2.44)
$\cos\theta^B_{\text{Sph}} < 0.9$	67.6	63.1	54.9
$\lvert\Delta E\rvert < 0.25$ $5.19 < m_{ES} < 5.3$	54.0 [30.5]	53.6 [28.8]	68.0 [56.0]
$\cos\theta^{\gamma}_{Hel}$	87.5 [88.7]	84.9 [85.9]	85.3 [86.2]
$0.01 < LAT_{\gamma} < 0.6$	90.1 [90.4]	89.2 [89.3]	85.2 [85.4]
$0.11 < M_{\pi^0} < 0.16$	94.7 [95.2]	94.2 [94.6]	92.1 [92.2]
$0.35 < \theta_{\gamma} < 2.39$	95.1 [95.3]	94.9 [95.1]	92.9 [93.0]
PID	32.6 [54.5]	28.0 [58.2]	22.9 [58.2]
$0.8 < m_{K^*} < 1.0\,\text{GeV}/c^2$ $0.55 < m_{\rho} < 0.95\,\text{GeV}/c^2$	6.4 [9.4]	7.5 [10.2]	18.4 [32.7]
$ANN > 0.2$	88.9 [89.4]	85.8 [88.3]	53.2 [48.0]
$-0.2 < \Delta E < 0.2$ $5.2 < m_{ES} < 5.3$ $ANN > 0.25$[> 0.4]	62.7 [50.2]	63.2 [49.4]	60.6 [38.6]
Total (Multiplicity)	(1.9 ± 0.1) $\times 10^{-4}$ (2.3 ± 0.1) $\times 10^{-4}$ (1.15)	(0.9 ± 0.1) $\times 10^{-4}$ (1.1 ± 0.1) $\times 10^{-4}$ (1.27)	(0.482 ± 0.002) $\times 10^{-2}$ (1.042 ± 0.002) $\times 10^{-2}$ (1.36)

Table 3.11: The efficiencies of the selection for MC simulated background samples. The $B^{\pm} \to \rho^{\pm}\pi^0$ analysis efficiencies are indicated in brackets and coloured red for clarity. The cuts are performed hierarchically (in the order given in the table), with the values being the efficiency of the cut with respect to the remaining sample of events.

Chapter 4

Analysis method and validation

4.1 Introduction

This chapter introduces the maximum likelihood fitting method and its implementation for this particular analysis. The structure of the probability density functions (PDFs) for the signal and background components is then given. The full treatment of backgrounds from B related sources which can fake signal events is described along with the expected number of events for each mode. Once the maximum likelihood fit has been described, the detailed validation studies that were undertaken to test its robustness and stability are explained. The chapter concludes with the details of the optimisation study for the ANN cut, which was performed only after proper confidence in the full fit model had been established.

4.2 The maximum likelihood fit

This section discusses how the method of maximum likelihood for parameter estimation is used to extract the signal yields and charge asymmetries from the data.

Suppose a general set of N observations of the variables $\mathbf{x} = [x_1, ..., x_n]$ is described by a PDF $\mathcal{P}(\alpha, \mathbf{x})$, where the underlying parameters $\alpha = [\alpha_1, ...\alpha_p]$ are unknown.

One may construct the likelihood $\mathcal{L}$ for the sample, with estimators $\hat{\alpha}$:

$$\mathcal{L}(\hat{\alpha}) = \prod_{i}^{N} \mathcal{P}(\hat{\alpha}, \mathbf{x}_i). \tag{4.1}$$

The likelihood is obviously a function of the parameter estimators $\hat{\alpha}$, and one can show that the choice of $\hat{\alpha}$ which maximises $\mathcal{L}$ provides an unbiased and efficient estimation of the true underlying parameters α [55].

One may extend the above formalism to the case where there are observations of multiple classes of event within the sample, and not only the functional form of the class's PDF but also the number of events observed from each class is unknown and hence to be determined. Of course the observed number of events, N_c, for class c itself follows a Poisson distribution with mean μ_c, and one might want to estimate the (unknown) *expected* number of events which is not necessarily equal to the observed number. In this case, the *extended maximum likelihood* formalism [55] provides a method to estimate the expected number of events:

$$\mathcal{L}(\hat{\mu}, \hat{\alpha})_c = \frac{\hat{\mu}_c^N e^{-\hat{\mu}}}{N_c!} \prod_{i}^{N_c} \mathcal{P}(\hat{\alpha}, \mathbf{x}_i). \tag{4.2}$$

It is usual to drop the factorial term in the denominator. The effective likelihood $\mathcal{L}'$ for all classes over the entire dataset is then given by

$$\mathcal{L}' = N_c!\ \mathcal{L} = (N')^N e^{-N'} \prod_{i=1}^{N} \mathcal{P}_i \tag{4.3}$$

where N' is the total number of expected events for all classes and N is the number seen. $\mathcal{P}_i$ is the total PDF which is normalised for the expected numbers of all sub-class components.

4.2.1 MINUIT and RooFit

The fit is ultimately handled by the software package MINUIT [56] which provides function minimisation and error estimation. The function which is minimised is the negative log-likelihood.

RooFit [57] is an object-oriented C++ toolkit developed at *BABAR* which interfaces to MINUIT and is fully integrated to the ROOT analysis framework [58]. RooFit handles the PDF construction, ensures proper normalisation over all variables and provides the front-end interface to the MINUIT minimisation routines. RooFit tools are used extensively in the analysis code.

4.3 Fit implementation

The unbinned extended maximum likelihood fit method is implemented for both analyses to extract the following parameters from the data.

- N^{Signal}: The total number of $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$ events in the case of the K^{*} analysis or the total number of $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ events in the case of the ρ analysis.
- N^{Cont}: The total number of continuum background events.
- $A^{Signal} = (N^{Signal(-)} - N^{Signal(+)})/N^{Signal}$: The charge asymmetry of the signal $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$ events or $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ events.
- A^{Cont}: The charge asymmetry of the continuum background events.

4.3.1 The event PDF and likelihood

The likelihood, $\mathcal{L}$, for the selected sample is given by the product of the PDFs for each individual event. The fit minimises the quantity $-2\ln\mathcal{L}$ which is equivalent to maximising $\mathcal{L}$. The PDF for a given event i is the sum of the signal and background terms, with an overall normalisation factor of $1/N$, N being the total number of events in the data sample. The likelihoods for each analysis are very similar, the only difference being the treatment of the mistag component for the $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ analysis. The PDFs used are described below.

$B^{\pm} \to K^{*\pm}\pi^0$ likelihood

The $B^{\pm} \to K^{*\pm}\pi^0$ event likelihood used is

$$\begin{aligned} N^{Total}\mathcal{P}_i &= N^{Signal} \cdot \frac{1}{2}\left[(1 - QA^{Signal}) \cdot f \cdot \mathcal{P}^{Signal}_{SCF,i} + (1 - QA^{Signal}) \cdot (1-f) \cdot \mathcal{P}^{Signal}_{True,i}\right] \\ &+ \sum_{j=1,n} N^{BB}_j \cdot \frac{1}{2} \cdot (1 - QA^{BB}_j) \cdot \mathcal{P}^{BB}_{j,i} \\ &+ N^{Cont} \cdot \frac{1}{2} \cdot (1 - QA^{Cont}) \cdot \mathcal{P}^{Cont}_i \end{aligned}$$

where P^x_i is the likelihood for mode x for event i, N^x is the expected number of events in that category, A^x is the charge asymmetry, Q is the charge of the kaon in the event and f is the fraction of SCF signal events. The latter is obtained from simulation and is fixed to 35.5%, as in Section 3.6. The total number of fitted events is therefore the sum of the true and self-cross-feed events.

$B^{\pm} \to \rho^{\pm}\pi^0$ likelihood

The $B^{\pm} \to \rho^{\pm}\pi^0$ event likelihood used is

$$\begin{aligned} N^{Total}\mathcal{P}_i &= \frac{N^{Signal}}{(1 + f_{Mistag})} \cdot \frac{1}{2}\Big[(1 - QA^{Signal}) \cdot f_{SCF} \cdot \mathcal{P}^{Signal}_{SCF,i} \\ &\quad + (1 - QA^{Signal}) \cdot (1 - f_{SCF}) \cdot \mathcal{P}^{Signal}_{True,i} \\ &\quad + (1 - QA^{Signal}) \cdot f_{Mistag} \cdot \mathcal{P}^{Signal}_{Mistag,i}\Big] \\ &+ \sum_{j=1,n} N^{BB}_j \cdot \frac{1}{2} \cdot (1 - QA^{BB}_j) \cdot \mathcal{P}^{BB}_{j,i} \\ &+ N^{Cont} \cdot \frac{1}{2} \cdot (1 - QA^{Cont}) \cdot \mathcal{P}^{Cont}_i \end{aligned}$$

where the terms are as above, f_{SCF} is the fraction of *correct charge* SCF signal events, and f_{Mistag} is the ratio of wrong charge to correct charge signal events. Thus the number of correct charge events (i.e.'True' + 'SCF') is equal to $N^{Signal}/(1+f_{Mistag})$, while the number of mistag events is equal to $f_{Mistag}.N^{Signal}/(1+f_{Mistag})$. Both fractions are obtained from MC simulation and are fixed to $f_{SCF} = 39.7\%$ and $f_{Mistag} = 7.6\%$, as described previously in Section 3.6. The total number of fitted signal events N^{Signal} is therefore the sum of the true, SCF and mistag events.

In both likelihood formulae the 4-D PDF for each mode is the product of the separate PDFs over the input variables, ΔE, m_{ES}, ANN_{Tran}, and $m_{K^*(\rho)}$:

$$\mathcal{P}_Q^k = \mathcal{P}_{m_{ES},\Delta E}^k \mathcal{P}_{ANN_{Tran}}^k \mathcal{P}_{\mathrm{mK}^*(\rho)}^k \tag{4.4}$$

where for continuum $\mathcal{P}_{m_{ES},\Delta E}$ is the product of two orthogonal one-dimensional parametric PDFs. However, due to correlations between ΔE and m_{ES} for signal and B–related background, $\mathcal{P}_{m_{ES},\Delta E}$ cannot be properly described by a product of two orthogonal, one-dimensional PDFs. In principle, one could form a parameterisation of ΔE in terms of m_{ES} (and vice-versa), but there is no general parameterisation available which describes all the various B backgrounds satisfactorily. Consequently we choose to describe the signal and B background PDFs $\mathcal{P}_{m_{ES},\Delta E}$ with a non-parametric two-dimensional PDF, called a 2-D *Keys* PDF. The *Keys* formalism is described in Section 4.3.2.

For $\mathcal{P}_{m_{K^*(\rho)}}^k$, again there is no generic parameterisation that works for all modes, so non-parametric one-dimensional PDFs (1-D *Keys*) are used to describe all modes except truth-matched signal, for which a Breit-Wigner and polynomial are used. The parameters for this signal PDF are held fixed in the main fit and then varied within their errors ($\pm 1\sigma$) in separate fits to estimate systematic uncertainties.

For the transformed neural network parameterisation, one-dimensional parametric forms are adopted for signal and B-related backgrounds. The parametric form is that of the 'Crystal Ball' lineshape defined in Eq.(4.6) below. The shape parameters are obtained from MC simulation and held fixed in the fit, but later varied within their errors in the systematic uncertainty studies. For continuum a 1-D *Keys* is used in the $B^\pm \to K^{*\pm}\pi^0$ analysis, but for the $B^\pm \to \rho^\pm\pi^0$ analysis a parametric form of an exponential+2^{nd} order polynomial is sufficient. The reason for the difference is due to the lower bound cut on ANN in the two cases; recall the $B^\pm \to K^{*\pm}\pi^0$ analysis has a lower bound cut of 0.25, the $B^\pm \to \rho^\pm\pi^0$ a cut of 0.4. The continuum distribution for ANN_{Tran} rises sharply for low values, and this rise is not well-described by the parametric form assumed for values above 0.4. Other parameterisations were investigated, but ultimately a 1-D *Keys* proved the most robust over the whole range of 0.25-1.00. In the case of the parametric description for the $B^\pm \to \rho^\pm\pi^0$ analysis, all the continuum ANN_{Tran} parameters are floated in the fit.

A brief description of the parametric forms assumed and the *Keys* method follows.

4.3.2 Description of sub-component PDFs

Keys

The *Keys* formalism is a kernel estimation method [59]. It is a non-parametric method for constructing a PDF, similar in principle to a histogram but invented to avoid the effect of binning. The principle is that each sample data *point* used in the estimation of the parent distribution is replaced by a *kernel* of probability about that point. In these analyses for the univariate case, the kernel is given by the Gaussian kernel $\hat{f}_i(x)$

$$\hat{f}_i(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(\frac{(x-t_i)^2}{2h^2}\right) \tag{4.5}$$

where t_i represents the data point and h is a parameter called the bandwidth. The total parent distribution is estimated by the sum over all kernels in the sample. For the multivariate case, the kernel is a product of univariate kernels. The choice of kernel is arbitrary and it is difficult to know the optimal choice for a general distribution. It is therefore reasonable to make the nominal choice by an educated inspection of the resultant PDF. This essentially means that care is taken to ensure the PDF is unlikely to be modelling statistical fluctuations. In this analysis the choice of bandwidth is made in this way for the nominal fit.

Crystal Ball lineshape

The Crystal Ball lineshape is essentially a Gaussian with a power-law tail that is used here for describing the transformed ANN variable. Its functional form is given by

$$\phi(x,\sigma,\alpha,n,\bar{x}) = \frac{1}{N}\begin{cases}\exp\left[-\frac{(x-\bar{x})^2}{2\sigma^2}\right], & x-\bar{x} > \alpha\sigma \\ \frac{\left(\frac{n}{\alpha}\right)^n \exp\left[-0.5\alpha^2\right]}{\left[\frac{n}{\alpha}-\alpha-\frac{(x-\bar{x})}{\sigma}\right]^n}, & x-\bar{x} \leq \alpha\sigma\end{cases} \tag{4.6}$$

ARGUS lineshape

The ARGUS lineshape [60] describes the continuum m_{ES} distribution. It is an empirically motivated parameterisation and takes the form

$$A(m_{ES}:m_0,\zeta) = \frac{1}{N} m_0 \sqrt{1-\left(\frac{m_{ES}}{m_0}\right)^2} \exp\left\{\zeta\left[1-\left(\frac{m_{ES}}{m_0}\right)^2\right]\right\} \tag{4.7}$$

where the kinematic cut-off $m_0 = 5.29\,\mathrm{GeV}/c^2$ is fixed to E^*_{beam}, N is an overall normalisation factor and ζ is the parameter that is floated.

4.4 PDF Projections

Projections of the PDFs for selected classes of event which illustrate the functional forms discussed above are shown in Figures 4.1 - 4.4.

4.5 B–related background

Events from other B decays provide a potentially difficult problem. From MC simulation it is expected that many of the channels that contribute to the B–related backgrounds will be ones that are poorly measured if at all.

The B background is treated as having two separate cases. Firstly there is the dominant contribution from $b \to c$ decays which is referred to here as the 'generic B background'. The second class of B backgrounds include charmless decays.

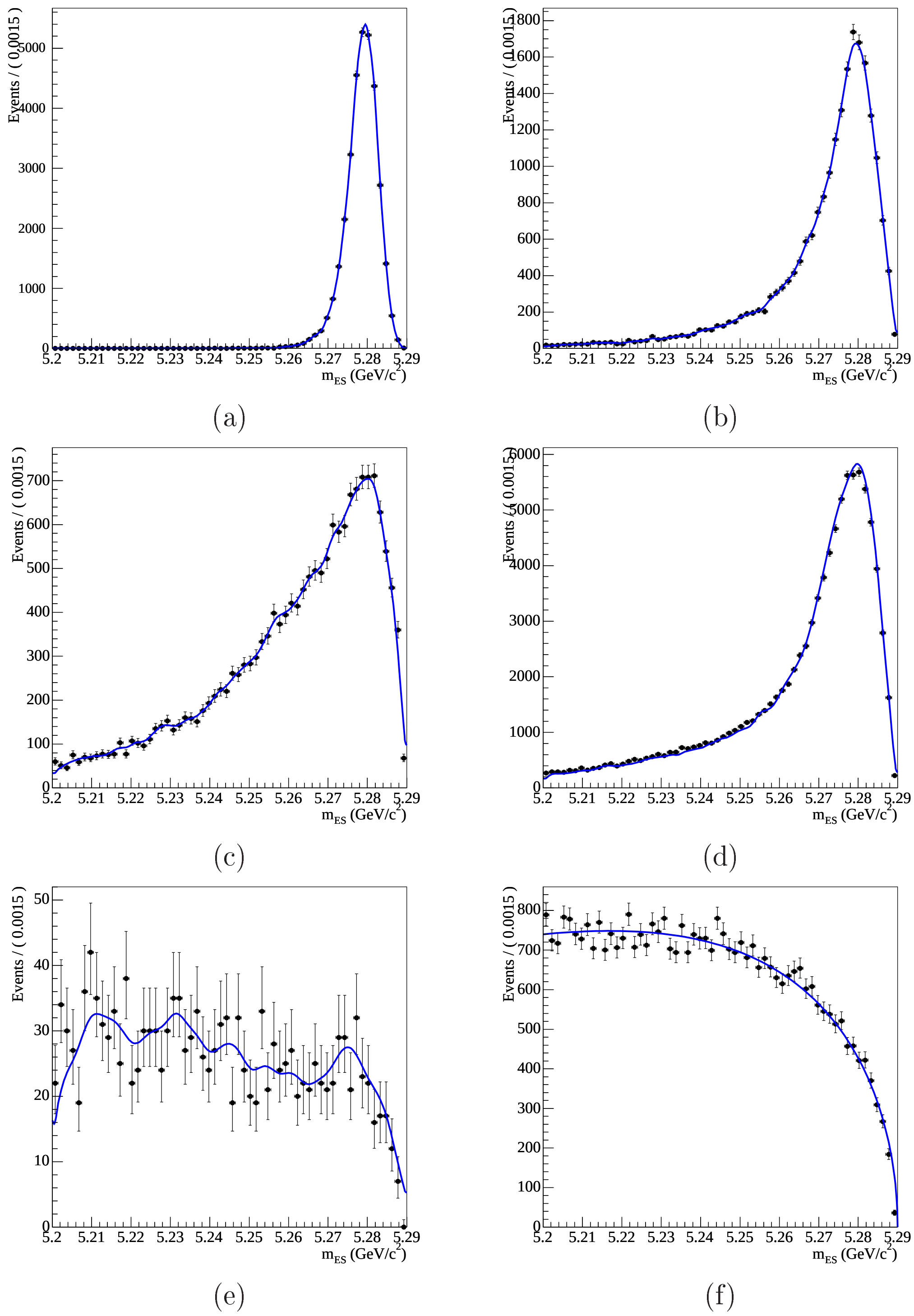


Figure 4.1: Examples of PDF projections for the m_{ES} distributions. The labels correspond to truth-matched $B^{\pm} \to \rho^{\pm}\pi^0$ (a), correct-charge SCF (b), mistag SCF (c), $B^0 \to \rho^+\rho^-$ (d), generic $b \to c$ (e) and continuum MC simulation (f).

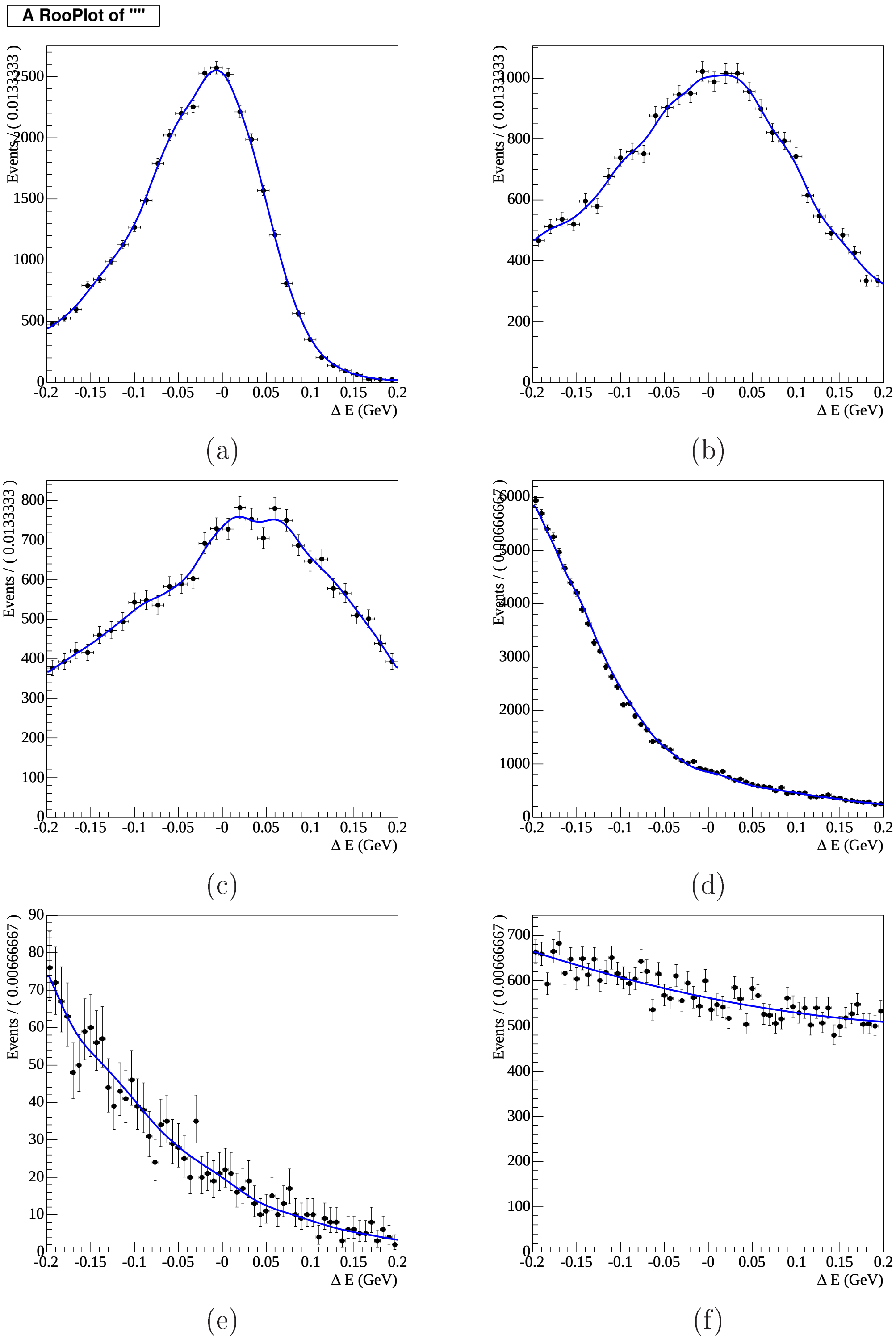


Figure 4.2: Examples of PDF projections for the ΔE distributions. The labels correspond to truth-matched $B^{\pm} \to \rho^{\pm}\pi^{0}$ (a), correct-charge SCF (b), mistag SCF (c), $B^{0} \to \rho^{+}\rho^{-}$ (d), generic $b \to c$ (e) and continuum MC simulation (f).

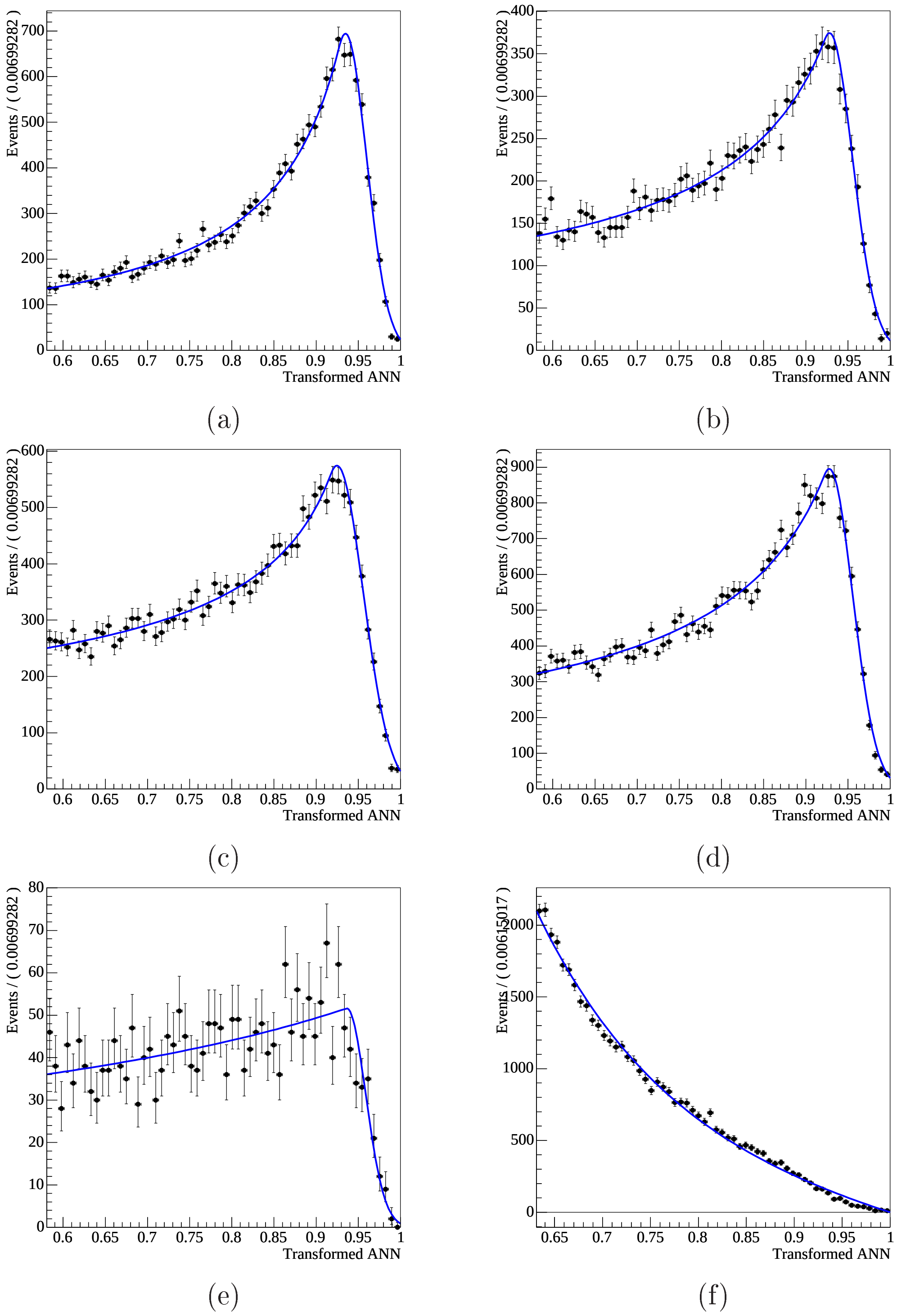
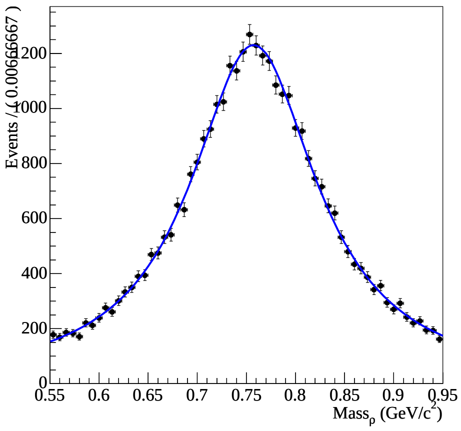
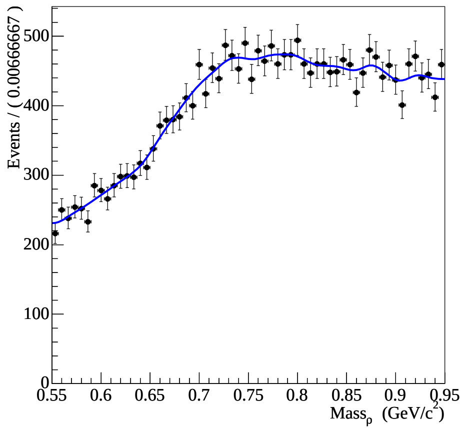


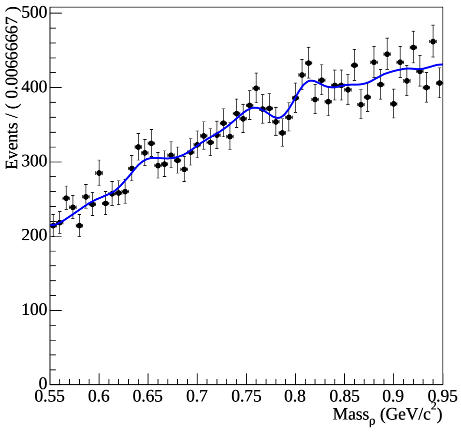
Figure 4.3: Examples of PDF projections for the ANN_{Tran} distributions. The labels correspond to truth-matched $B^{\pm} \to \rho^{\pm}\pi^{0}$ (a), correct-charge SCF (b), mistag SCF (c), $B^{0} \to \rho^{+}\rho^{-}$ (d), generic $b \to c$ (e) and continuum MC simulation (f).



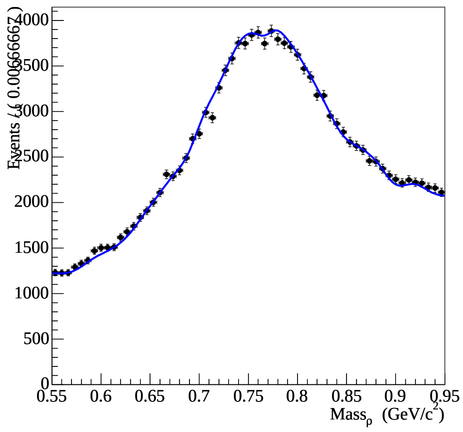
(a)



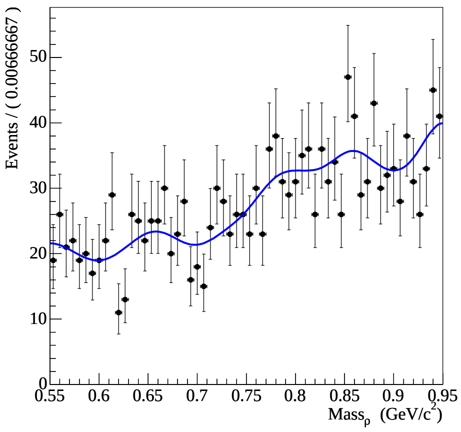
(b)



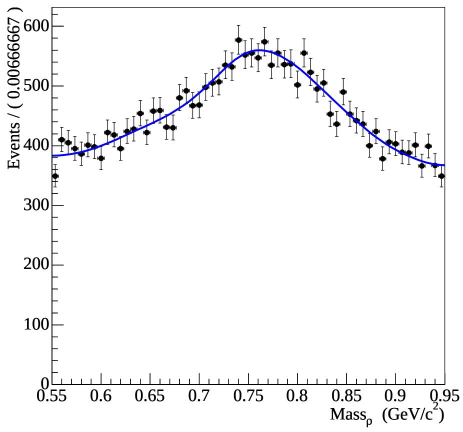
(c)



(d)



(e)



(f)

Figure 4.4: Examples of PDF projections for the $\pi^{\pm}\pi^{0}$ mass distributions. The labels correspond to truth-matched $B^{\pm} \to \rho^{\pm}\pi^{0}$ (a), correct-charge SCF (b), mistag SCF (c), $B^{0} \to \rho^{+}\rho^{-}$ (d), generic $b \to c$ (e) and continuum MC simulation (f).

To account for the first background case, generic $B\overline{B}$ MC events are studied to identify the events where both the B mesons decay inclusively to charm. This information is taken from the MC truth and it is guaranteed to remove not only the signal modes from the sample (which will be present in the generic sample at some nominal level), but also the second type of background which can be studied in isolation. The generic B sample includes events coming from both B^+B^- and $B^0\overline{B}^0$ events and these are incorporated into one component in the fit. The PDFs for this component are shown in Section 4.4.

The second class of background tends to be more problematic as the branching fractions can be poorly known, and because potentially they can peak at the same places as the signal events. Over 30 individual MC modes were studied which have the potential to fake the signal. These modes were identified from the MC truth for charmless events that were present in the generic B sample before the non-charm veto was applied, and also from anticipating modes that were not listed in the MC decay files but could in principle be in the data at some level (typically with branching fractions estimated to be in the range $10^{-5} - 10^{-6}$). The modes considered are fully detailed in Section 4.5.4. These modes are added into the fit at the levels shown in the tables with the expected number entering the fit fixed. The modes with measured branching fractions and CP asymmetries are assumed to have the branching fraction and asymmetries as quoted by the Heavy Flavour Averaging Group (HFAG) [17] unless otherwise indicated. For the modes where only an upper limit is known, the expected number is estimated by assuming a branching fraction of half the referenced limit, with an assigned error of 100% of this value. As part of the systematic uncertainty studies the numbers and asymmetries are varied by the measured or theoretical error on the relative branching fractions.

Some of the modes bear comment. Although the modes $B^0 \to \rho^{\pm}\ \rho^{\mp}$ and $B^{\pm} \to \pi^0\pi^0$may in principle display direct CP asymmetries, they will not affect the analysis in as far as manifesting an artificial charge asymmetry in reconstructing signal events is concerned. Thus their asymmetry is set to zero and not varied as a systematic. The radiative inclusive $X_s\gamma$ background implicitly includes the mode $K^*\gamma$. The $K^*\gamma$

background is significant enough to merit study in its own right. When considering $X_s\gamma$ in the fit therefore, the $K^*\gamma$ contribution is removed from $X_s\gamma$ MC, and the assumed BF for this mode is adjusted to reflect the fact that the $K^*\gamma$ is treated separately.

4.5.1 Higher $K^{**\pm}$, $\rho^{*\pm}$ and non-resonant backgrounds

The possible presence of non-resonant $B^\pm \to K^\pm\pi^0\pi^0$ ($B^\pm \to \pi^\pm\pi^0\pi^0$) decays as well as $K^{**\pm}$ ($\rho^{*\pm}$) resonances[1] can affect the interpretation of the $B^\pm \to K^{*\pm}\pi^0$ ($B^\pm \to \rho^\pm\pi^0$) branching fraction and asymmetry in the following ways:

- Non–resonant $B^\pm \to K^\pm\pi^0\pi^0$ ($B^\pm \to \pi^\pm\pi^0\pi^0$) amplitudes can interfere with the $B^\pm \to K^{*\pm}\pi^0$ ($B^\pm \to \rho^\pm\pi^0$) amplitudes in the K^* ($\rho^\pm$) mass region. If this interference is significant – either constructively or destructively – then the interpretation of the $\mathcal{B}$ and $\mathcal{A}_{CP}$ is uncertain.
- Amplitudes for $B^\pm \to K^\pm\pi^0\pi^0$ ($\pi^\pm\pi^0\pi^0$) decays via $K^{**\pm}$ ($\rho^{*\pm}$) resonance states can also interfere with the $B^\pm \to K^{*\pm}\pi^0$ amplitudes in Dalitz regions where the K^* (ρ) and $K^{**\pm}$ ($\rho^{*\pm}$) masses cross.

The possible quantum mechanical interference from these modes can be problematic if there is a significant amplitude to them. Experimentally, these interference effects can only be determined from a complete analysis of the Dalitz plane and fits to the amplitude contributions, however this is beyond the scope of this analysis. An alternative strategy to estimate the possible contributions is detailed in the next section below.

Decays of higher $K^{*\pm}$ type ($\rho^\pm$-type) resonances where the final state is *not* $K^\pm\pi^0\pi^0$ ($\pi^\pm\pi^0\pi^0$) can be treated as B–related background, and there is no interference for these modes. The polluting contribution they may give can be estimated from simulation.

[1] K^{**+} is used here to denote excitations of the $u\bar{s}$ bound state that have masses above the $K^{*\pm}$ (892) and that decay to the $K^+\pi^0$ final state. Similarly ρ^{*+} is used to denote excitations of the $u\bar{d}$ state above the $\rho(770)$ that decay to $\pi^+\pi^0$ [15].

4.5.2 Effect of possible interfering amplitudes on the branching fraction measurements from non–resonant decays

The non–resonant $B^{\pm} \to K^{\pm}\pi^{0}\pi^{0}$ and $B^{\pm} \to \pi^{\pm}\pi^{0}\pi^{0}$ branching fractions have, to date, not been measured. Although the resonant mass cuts both exclude a large portion of the non–resonant phase space ($\sim 96\%$ for $B^{\pm} \to K^{*\pm}\pi^{0}$ and $\sim 91\%$ for $B^{\pm} \to \rho^{\pm}\pi^{0}$)[2], any presence of this mode will inevitably give a contribution to our signal since it looks similiar in the ΔE, m_{ES} and ANN distributions (though will be flatter in the resonant mass region). To estimate the significance of its contribution we look into a region of the Dalitz plot which is far from our signal and from the potential $K^{**\pm}$ and or ρ resonances discussed below, and which has relatively small amounts of continuum pollution. Any non–resonant signal is then extrapolated into the resonant region.

The PDFs used in the non–resonant fit are all assembled from MC samples for the modes expected in the region. In both cases the only significant backgrounds expected are from generic B and continuum events. The yields of the generic B background are fixed to values expected from MC simulation and PDF shape parameters fixed to the values obtained from fitting the MC samples. The continuum background yield is left a floating parameter.

This strategy is similar to other quasi-two body $\rho\pi$ analyses at *BABAR* [61] and assumes that non–resonant production is uniformly populated over the whole Dalitz plane.

Non–Resonant $K^{\pm}\pi^{0}\pi^{0}$

As no $K^{*\pm}$ information is implicitly present in our best candidate selection, neural network discriminating variable or preselection, we can use the events that are rejected by our $K^{*\pm}$ mass cut and which lie in the suitable region of the Dalitz phase

[2] These numbers are obtained from the ratio of the area occupied by the resonant band in the Dalitz plane to the area of the whole Dalitz plane, assuming the non–resonant mode populates the Dalitz plane uniformly.

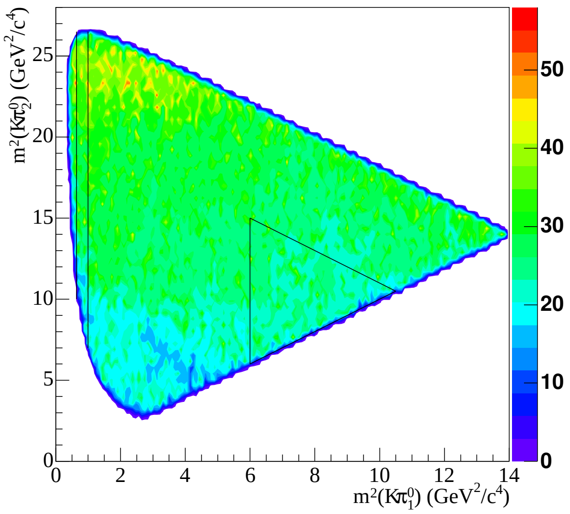


Figure 4.5: Dalitz plot showing K^* mass region (left hand strip) and region from which the number of non–resonant decays is estimated (triangular region), referred to as *region A* in the text.

space (hereafter region A, see Figure 4.5) without additional treatment. In a similar way as for the $K^{*\pm}$ signal, PDFs for B backgrounds, continuum and non–resonant signal are formed from MC simulation for the three discriminating variables ΔE, m_{ES} and ANN_{Tran}, for events in region A. A likelihood fit is then performed in the same phase space region on data using these PDFs to extract the non–resonant yield in region A, N_{NR}. In this case, due to the low expected yields, there is no fit for the charge asymmetry and the candidate charge does not enter into the fit.

Once obtained, N_{NR}, is then used to conservatively estimate the yield of non–resonant $B^{\pm} \to K^{\pm}\pi^0\pi^0$ in the $K^{*\pm}$ band with the formula $N_{NR}^{K^{*\pm}} = R \times N_{NR}$. The coefficient R, equal to 0.41, is obtained from a sample of 700k non–resonant $B^{\pm} \to K^{\pm}\pi^0\pi^0$ MC simulated events and is taken as the ratio of the number of reconstructed non–resonant events in the $K^{*\pm}$ region to the number of reconstructed non–resonant events in region A. This implicitly takes into account variations in the efficiency across the Dalitz plane.

A likelihood fit in region A yields 6.3 ± 7.9 non-resonant events. Since we consequently expect less than 3 events in the K^* signal region the non-resonant contribution is deemed negligible, along with any interference due to it.

Non–Resonant $\pi^{\pm}\pi^0\pi^0$

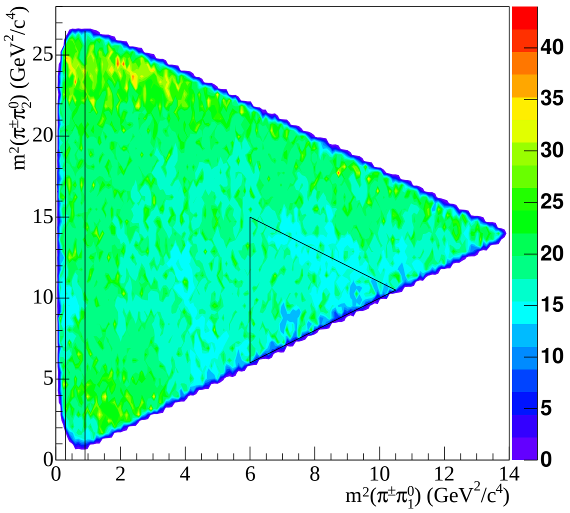


Figure 4.6: Dalitz plot showing $\rho^{\pm}$ mass region (left hand strip) and region from which the number of non–resonant decays is estimated (triangular region), referred to as *region B* in the text.

In an analagous way to the non–resonant $K^{\pm}\pi^0\pi^0$, the non–resonant $\pi^{\pm}\pi^0\pi^0$ is estimated from a fit to a sideband in the Dalitz plot. The sideband is shown graphically in Figure 4.6. There is one principal 'difference' to the $K^{\pm}\pi^0\pi^0$ estimation, namely that the best candidate selection for the non–resonant $\pi^{\pm}\pi^0\pi^0$ study is the *same* as for the $K^{\pm}\pi^0\pi^0$, i.e. a selection based upon a χ^2 formed from the π^0 masses. This is because it makes no sense to apply the best candidate selection which is based on the ρ helicity to the Dalitz sideband. However, it should be emphasized there is actually no difference to the overall extrapolated number of events by swapping between the algorithms because the best candidate selection by definition never throws away events. The only difference the selection algorithm makes is which candidate is actually chosen, which can change the shapes of the PDFs as they are in general correlated with the Dalitz variables. Thus it is important for any non–resonant PDF made for the actual ρ fit to be made with the proper algorithm, i.e. the helicity based selection. The extrapolation ratio is 0.80.

A likelihood fit in region B yields -5.1 ± 7.6 non–resonant events in a data sample of 1100 events. This value, though negative, is consistent with zero. Clearly extrapolating this number into the ρ signal region is not physically meaningful, and therefore the non–resonant contribution is also treated as being negligible along with associated interference effects. The only main background contributions in the non–resonant region selected are continuum and generic B events, the latter of which is expected to be at the level of around 50 events.

4.5.3 Estimation of higher resonance pollution

Higher $K^{**\pm}$ pollution

Some of the higher $K^{**\pm}$ resonances can decay to a $K^{\pm}\pi^0$ final state (Table 4.1). Although all $K^{**\pm}$ states peak well outside our $K^*(892)$ mass cut, some may still feed in to the selection due to their large width.

State	Mass [MeV/c^2]	Width [MeV/c^2]
$K^*(892)$	891.66 ± 0.236	50.8 ± 0.9
$K^*(1410)$	1414 ± 15	232 ± 21
$K_0^*(1430)$	1412 ± 6	294 ± 23
$K_2^*(1430)$	1425.6 ± 1.5	98.5 ± 2.9
$K^*(1680)$	1717 ± 27	322 ± 110
$K_3^*(1780)$	1776 ± 7	159 ± 21
K^{*0} (1950)	$1945 \pm 10 \pm 20$	$201 \pm 34 \pm 79$
$K_4^*(2045)$	2045 ± 9	198 ± 30
$K_5^*(2380)$	$2382 \pm 14 \pm 19$	$178 \pm 37 \pm 32$

Table 4.1: Possible $K^{**\pm}$ resonances decaying to $K\pi$ final states.

Conservative estimates for the contribution of these higher modes to our signal are made. To estimate the contribution a cut is made in the $K^{\pm}\pi^0$ invariant mass between 1.2 and 1.6 GeV/c^2; this selection is motivated by the presence of the broad $K^{*\pm}_0(1430)$ resonance which is expected to be the dominant source of pollution due to its large width (0.3 GeV/c^2 [15]). The $K^{*\pm}_0(1430)$ is a $J^P = 0^-$ resonance, and therefore any interference with the vector $K^*(892)$ should integrate to zero over the full $\cos\theta_{Hel}$ range (as described in Chapter 1), providing the distribution in $\cos\theta_{Hel}$ is symmetric. However, due to selection efficiencies this is not the case, and the effect of interference may actually be observed. To be conservative in accounting for possible interference effects, a 100% error is assigned to the extrapolated number. The $K^{*\pm}(1410)$ has a small branching fraction $(6.6 \pm 1.3)\%)$ [15] to the $K^\pm\pi^0$ final state, and therefore is not thought to make a significant contribution to any yield seen in the sideband. If it is present in some small quantity, any interference it may have with the $K^*(892)$ in its tail should be accounted for by the large systematic assigned.

In this region we make another full maximum likelihood fit to the three variables ΔE, m_{ES} and ANN in an analogous way to the signal analysis. PDFs are formed for continuum, B backgrounds and signal. B backgrounds are fixed, continuum yields and shapes are floated. Studies indicated there were no significant biases observed with these PDFs.

After fitting the yield of $K^{**\pm}$ from the sideband, we use $B^\pm \to K^{*\pm}(1430)\pi^0$ MC events to extrapolate the numbers expected into the signal region. In the MC sample, the ratio of candidates in the sideband to candidates in the signal region is approximately 7.7:1. The fit in the sideband yields 261 ± 34 events, giving an estimate of the $K^{**\pm}$ pollution as 34 ± 34 events.

The effect of the choice of range for the $K^{**\pm}$ invariant mass was investigated by varying the range and refitting to the data. In all cases the PDFs for every mode falling in the sideband were remade to account for the change in selection, along with their associated yields. The results of these fits were compared to those expected from the $B^\pm \to K^{*\pm}(1430)\pi^0$ MC simulation. The results of this investigation are shown in Table 4.2. The agreement between data and MC is close enough within errors to believe in the presence of real $K^{**\pm}$ resonances in the sideband.

The $K^{**\pm}$ pollution enters as a background component in the nominal fit. The PDFs used to model it are composed from MC simulated $B^\pm \to K^{*\pm}(1430)\pi^0$ events.

Mass range (GeV/c^2)	Yield from fit	Expected yield from MC sim.
1.25-1.55	217 ± 30	180
1.20-1.60	261 ± 34	261
1.15-1.65	301 ± 41	284

Table 4.2: Summary of the effect of the choice of $K^{**\pm}$ mass range. The third column indicates the yield expected from $B^\pm \to K^{*\pm}(1430)\pi^0$ assuming a yield of 261 events in the chosen range of 1.20-1.60 GeV/c^2.

Higher ρ^* pollution

Similarly for the $B^\pm \to \rho^\pm\pi^0$ analysis, there are potential higher rho-type resonances which can enter into the signal mass region and possibly interfere. The branching fractions of these $B^\pm \to \rho^{*\pm}\pi^0$ decays are, to date, unmeasured. The only non-strange vector resonances which can decay to two pions are the $\rho(1450)$ and $\rho(1700)$. These resonances are often considered to be higher radial excitations of the $\rho(770)$ [62]. To account for the possible presence of these modes, a fit to the $B^\pm \to \rho^{*\pm}\pi^0$ yield is performed in a sideband of the invariant mass using the three variables m_{ES}, ΔE and ANN_{Tran}. The mass window was chosen to be $1.5 < m_\rho < 2.0\,\mathrm{GeV}/c^2$, so as to be centred near the pole of the $\rho(1700)$, and to be as far from the $\rho(770)$ as possible. In this case, unlike the $B^\pm \to K^{*\pm}\pi^0$ analysis, the $B^\pm \to \rho^\pm\pi^0$ branching fraction is already measured [17] and the fit in the sideband includes this mode as a background. The size of the expected background yield from the $\rho(770)$ (and hence the normalisation of its fixed contribution) is about 41 events from MC simulation, which are mostly of the SCF type.

The fitted yield for the $B^\pm \to \rho^{*\pm}\pi^0$ decay, is then extrapolated into the nominal region. However to be conservative, although the choice of mass range is motivated by the $\rho(1700)$ any yield seen is attributed entirely to the $\rho(1450)$, which is the closer of the two resonances. The extrapolation is then assigned the conservative systematic of 100%, as in the case of the $B^\pm \to K^{*\pm}\pi^0$ analysis. From the $B^\pm \to \rho^\pm(1450)\pi^0$ MC, the ratio of candidates in the sideband to candidates in the signal mass region is approximately 12.6:1. The fit in the sideband yields 101 ± 32 events, giving an estimate of the ρ^* pollution as 8 ± 8 events. The higher $\rho^{*\pm}$ then enters into the nominal fit with PDFs constructed from $B^\pm \to \rho^\pm(1450)\pi^0$ MC simulation.

Mass range (GeV/c^2)	Yield from fit	Expected yield from MC sim.
1.45-2.05	128 ± 36	137
1.5-2.00	101 ± 32	101
1.55-1.95	45 ± 27	70

Table 4.3: Summary of the effect of the choice of ρ^* mass range. The third column indicates the yield expected from $B^\pm \to \rho^\pm(1450)\pi^0$ MC simulation assuming a yield of 101 events in the chosen range of 1.50-2.00 GeV/c^2.

The effect of the choice of mass range was also investigated. Two additional fits were performed in the ranges $1.45 - 2.05\,\mathrm{GeV}/c^2$ and $1.55 - 1.95\,\mathrm{GeV}/c^2$ and the results were compared with the levels expected from $B^\pm \to \rho^\pm(1450)\pi^0$ MC simulation. This study is summarised in Table 4.3. As can be seen, the fits are compatible with the levels expected from MC simulation, giving justification for extrapolating the yield of 101 events as though they were entirely from the decay $B^\pm \to \rho^\pm(1450)\pi^0$.

4.5.4 Summary of B backgrounds

Tables 4.4 and 4.5 show the expected levels of B-related backgrounds for the $B^\pm \to K^{*\pm}\pi^0$ and $B^\pm \to \rho^\pm\pi^0$ analyses respectively, along with their associated error.

Backgrounds to $K^{*\pm}\pi^0$ from B-related Sources				
Mode	Efficiency (%)	Assumed $\mathcal{B}$ ($\times10^{-6}$)	Assumed $\mathcal{A}_{CP}$	Expected yield
(1) Generic $b \to c$	$1.35\ \times10^{-4}$	1.0	0.00 ± 0.00	310.3 ± 17.6
(2) $B^{\pm} \to K^{\pm}\pi^0$	3.63	12.1 ± 0.8 [17]	0.04 ± 0.04 [17]	101.6 ± 6.7
(3) $B^0 \to K^{\pm}\rho^{\mp}$	2.14	9.9 ± 1.6 [17]	0.17 ± 0.16 [17]	48.9 ± 7.9
(4) $B^0 \to K^{*\pm}\rho^{\mp}$	1.17	12.0 ± 12.0 [17]	0.00 ± 0.20	32.4 ± 32.4
(5) $B^{\pm} \to \rho^{\pm}\rho^0$	0.40	26.4 ± 6.4 [17]	0.00 ± 0.20	24.4 ± 5.5
(6) $B^{\pm} \to K^{*\pm}\gamma$	0.27	40.2 ± 2.6 [17]	-0.01 ± 0.03 [17]	25.1 ± 2.2
(7) $B^{\pm} \to \rho^{\pm}\pi^0$	0.70	12.0 ± 2.0 [17]	0.16 ± 0.13 [17]	19.5 ± 3.2
(9) $B^{\pm} \to X_s\gamma$	0.02	315 ± 30 [15]	0.00 ± 0.20	12.4 ± 1.2
(9) $B^0 \to \rho^{\pm}\pi^{\mp}$	0.22	24.0 ± 2.5 [17]	0.00 ± 0.20	12.2 ± 1.3
(10) $B^0 \to \pi^0 K^{*0}$	3.12	1.7 ± 0.8 [17]	0.00 ± 0.20	12.2 ± 5.8
(11) $B^0 \to \rho^{\pm}\rho^{\mp}$	0.15	30.0 ± 6.0 [17]	0.00 ± 0.00	10.4 ± 2.1
(12) $B^{\pm} \to \rho^{\pm}K^{*0}$	0.40	9.2 ± 2 [17]	-0.14 ± 0.20 [17]	8.5 ± 1.8
(13) $B^{\pm} \to a_1^{\pm}\pi^0$	0.12	17.5 ± 17.5 [63]	0.00 ± 0.20	4.8 ± 4.8
(14) $B^{\pm} \to \pi^{\pm}\pi^0$	0.28	5.5 ± 0.6 [17]	-0.02 ± 0.07 [17]	3.6 ± 0.4
(15) $B^{\pm} \to K^{**}\pi^0$	-	-	0.00 ± 0.20	33.9 ± 33.9
Total				660 ± 40

Table 4.4: Summary of B backgrounds to $B^{\pm} \to K^{*\pm}\pi^0$.

Backgrounds to $\rho^\pm\pi^0$ from B-related Sources				
Mode	Efficiency (%)	Assumed $\mathcal{B}$ ($\times 10^{-6}$)	Assumed $\mathcal{A}_{CP}$	Expected yield
(1) Generic $b \to c$	1.69×10^{-4}	-	0.00	392.3 ± 19.8
(2) $B^0 \to \rho^\pm\rho^\mp$	2.12	30.0 ± 6.0 [17]	0.00 ± 0.00	147.7 ± 29.5
(3) $B^0 \to \rho^\pm\pi^\mp$	2.39	24 ± 2.5 [17]	0.00 ± 0.20	132.9 ± 13.8
(4) $B^\pm \to X_s\gamma$	0.08	$315 \pm \pm 30$ [17]	0.00 ± 0.04 [17]	57.6 ± 5.5
(5) $B^\pm \to \pi^\pm\pi^0$	4.12	5.5 ± 0.6 [17]	-0.02 ± 0.07 [17]	52.5 ± 5.7
(6) $B^0 \to a_1\pi^0$	1.16	17.5 ± 17.5 [63]	0.00 ± 0.20	46.9 ± 46.9
(7) $B^\pm \to \rho^\pm\rho^0$	0.57	26.4 ± 6.4 [17]	-0.09 ± 0.16 [17]	34.6 ± 8.4
(8) $B^\pm \to \pi^\pm K_s(\to \pi^0\pi^0)$	1.49	3.74 ± 0.20 [17]	-0.02 ± 0.03 [17]	12.9 ± 0.7
(9) $B^\pm \to K^\pm\pi^0$	0.41	12.1 ± 0.8 [17]	0.04 ± 0.04 [17]	11.5 ± 0.8
(10) $B^0 \to \pi^0\pi^0$	2.62	1.51 ± 0.28 [17]	0.00 ± 0.00	9.2 ± 1.7
(11) $B^0 \to \eta\prime\pi^0$	1.15	1.85 ± 1.85 [17]	0.00 ± 0.20	4.9 ± 4.9
(12) $B^\pm \to \pi^\pm K^{*0}$	0.17	9.76 ± 1.22 [17]	0.00 ± 0.20	3.8 ± 0.5
(13) $B^0 \to \pi^0 K^{*0}$	0.90	1.7 ± 0.8 [17]	0.00 ± 0.20	3.5 ± 1.7
(14) $B^\pm \to K^{*\pm}\pi^0$	0.61	2.3 ± 0.7 [64]	0.00 ± 0.20	3.4 ± 1
(15) $B^0 \to K^\pm\rho^\mp$	0.14	9.9 ± 1.6 [17]	0.17 ± 0.16 [17]	3.2 ± 0.5
(16) $B^\pm \to K^{*\pm}\gamma$	0.02	40.3 ± 2.6 [17]	-0.01 ± 0.03 [17]	2.1 ± 0.1
(17) $B^\pm \to \rho^\pm\gamma$	0.65	0.9 ± 0.9 [17]	0.00 ± 0.20	1.4 ± 1.4
(18) $B^0 \to K^{*\pm}\rho^\mp$	0.05	12.0 ± 12.0 [17]	0.00 ± 0.20	1.4 ± 1.4
(19) $B^\pm \to \rho^\pm(1450)\pi^0$	-	-	0.00 ± 0.20	8 ± 8

Table 4.5: Summary of B backgrounds to $B^\pm \to \rho^\pm\pi^0$.

4.6 Fit validation

The complete fit model has now been fully described, with due consideration given to all the significant backgrounds that affect each analysis. However, this in no way justifies its application to the data, rather it must be proven to be robust to a variety of differing validation tests. This is the subject of the remainder of the chapter.

4.6.1 Fits to individual modes

The first and simplest check to validate the integrity of the fitting code is to set all the B backgrounds in the likelihood to zero by hand, and then apply the fit to two individual MC samples – continuum and then signal – but floating the continuum and signal yields as intended in the final fit. One would expect the fit to return the true amounts of both, within a reasonable error. The results of these fits are shown in Table 4.6. The fitted results are in agreement with the expected values, providing a positive initial check of our fit. The perfect agreement of the fitted number with the actual number generated given the error of 449 is merely down to the fit in this case returning the seeded value which is the number of generated signal events. Different initial starting values return differing fitted yields consistent with the size of this error. The large error from fitting the continuum to the signal MC (49 ± 1400) are due to the problems inherent in fitting the parameters of the continuum PDFs to what is essentially nothing.

A test was also conducted whereby a fit was performed to a sample composed of continuum, generic B and signal MC. The results of this fit can be seen in Table 4.7. Again the fitted results are in agreement with the initial sample sizes.

4.6.2 Toy Monte Carlo studies

It is important to validate that the fit is performing properly, i.e. that it is not biased and that the errors on the fit yields are estimated correctly. To initially test

Fitted Mode		Generated No. of Events [Asymmetry]	
		$B^{\pm} \to K^{*\pm}\pi^0$	Continuum
Generated number		201522 [0.0]	9862 [0.0]
$B^{\pm} \to K^{*\pm}\pi^0$	N^{Fit}	201522 ± 449	0.81±1.26
	A^{Fit}	−0.01±0.016	2.9±3.0
Continuum	N^{Fit}	49±1400	9861.1 ± 99.7
	A^{Fit}	0.38 ± 4.00	−0.042± 0.012

Table 4.6: Results of the full fit applied to various MC samples. The fitted parameters are all in agreement with expectation.

Fitted Mode	Generated No. of Events [Asymmetry]		
	$B^{\pm} \to K^{*\pm}\pi^0$	Continuum	Generic B $(b \to c)$
Input number	10000 [0.0]	9862 [0.0]	662 [0.0]
N^{Fit}	10042 ± 111	9970±132	513±88
A^{Fit}	−0.015±0.011	−0.040±0.012	−0.036± 0.17

Table 4.7: Results of the full fit applied to a single MC samples composed of generic $B(b \to c)$, continuum and signal. The fitted parameters are all in agreement with expectation, within errors.

the fit, 'toy' MC studies were performed. These involved 1000 separate experiments where each experiment involved using the PDFs to generate a dataset of 'events' with values for the four variables $[m_{ES}, \Delta E, ANN_{Tran}, m_{\rho,K^{*\pm}}]$. Each dataset is then treated like the real one, with the full fit performed to extract both the signal and continuum yields and asymmetries. The generated number for each fit component was determined by randomly fluctuating the expected number according to a Poisson distribution. The expected numbers for each mode are taken from Tables 4.4 and 4.5 for the B backgrounds, along with ~ 22000 (~ 47000) continuum and ~ 110 (~ 380) signal events, the latter being the expected number of K^* (ρ) events for a branching fraction of 3 (12)$\times 10^{-6}$. In addition, samples were created with differing asymmetry values, $\mathcal{A}_{CP} = -1.0, \ -0.5, \ 0.0, \ 0.5, \ 1.0$.

The potential bias within the fit is examined by forming the pull which is defined as:

$$\text{Pull} \;=\; \frac{X^j_{Fit} - X^j_{Exp}}{\sigma^j_{Fit}} \tag{4.8}$$

Here $X^j_{Fit(Exp)}$ can be either the number of events or the charge asymmetry fitted (expected) for mode j with σ^j_{Fit} the error on the fit variable.

Figure 4.7 shows the pull distributions for the $B^{\pm} \to K^{*\pm}\pi^0$ mode. Pulls for the $B^{\pm} \to \rho^{\pm}\pi^0$ studies are shown in Figure 4.8. In both cases the pulls are centred on zero as one might expect and the widths are approximately one. This gives confidence that the fit is structured correctly and performing as expected. In the case of the asymmetry pulls, it can be seen that the fit slightly overestimates the error. The exact causes of this are not known as the fit error estimation is non–trivial.

However, a solution is simply to multiply the error on any final measurement on data by the appropriate correction factor – after all this is one of the purposes of the toy experiments. Further discussion of this point is left until the final chapter.

4.6.3 Toy studies using full Monte Carlo simulation

As a further validation check yet another series of toy Monte Carlo studies were undertaken where, for certain components including the signal, fully simulated Monte Carlo events were used instead of generating with the PDFs. The selection of these fully simulated events was done randomly and care was taken to ensure there were ample statistics available in the component samples such that repeated selection of events from experiment to experiment was rare. All the samples used were events which had passed the selection criteria. These studies were used to test primarily that the signal component is correctly modelled in the fit, taking into account any subtle correlations that might not be obvious initially.

The results of these studies are shown in Figures 4.9 and 4.10. There are no biases seen in the number of fitted signal events nor in the asymmetry. This gives further

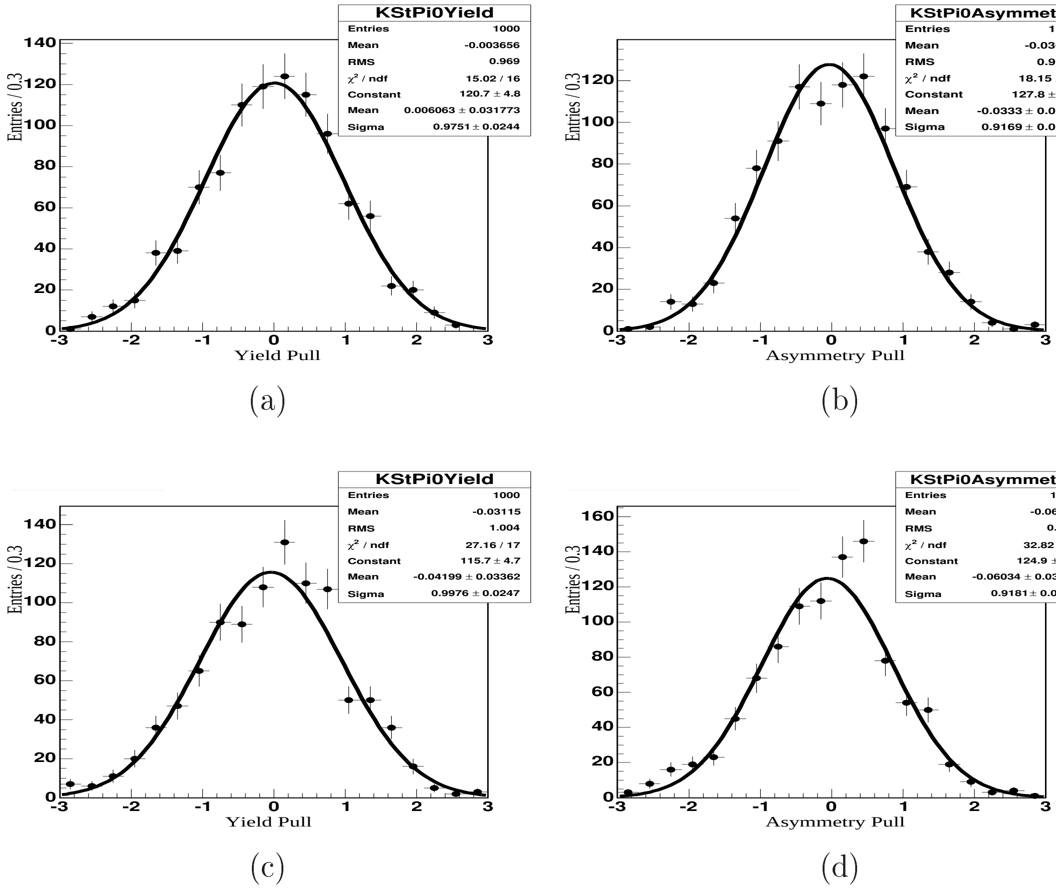


Figure 4.7: $B^{\pm} \to K^{*\pm}\pi^0$ pull distributions from toy MC studies. The plots show the pulls for the total number of signal (left) and $\mathcal{A}_{CP}$ (right), Superimposed are Gaussian fits to the data. (a) and (b) are for $\mathcal{A}_{CP}$ =0, while (c) and (d) are for $\mathcal{A}_{CP}$ =0.5. Pull distributions for the other non–zero $\mathcal{A}_{CP}$ values are also consistent with zero bias and unit width.

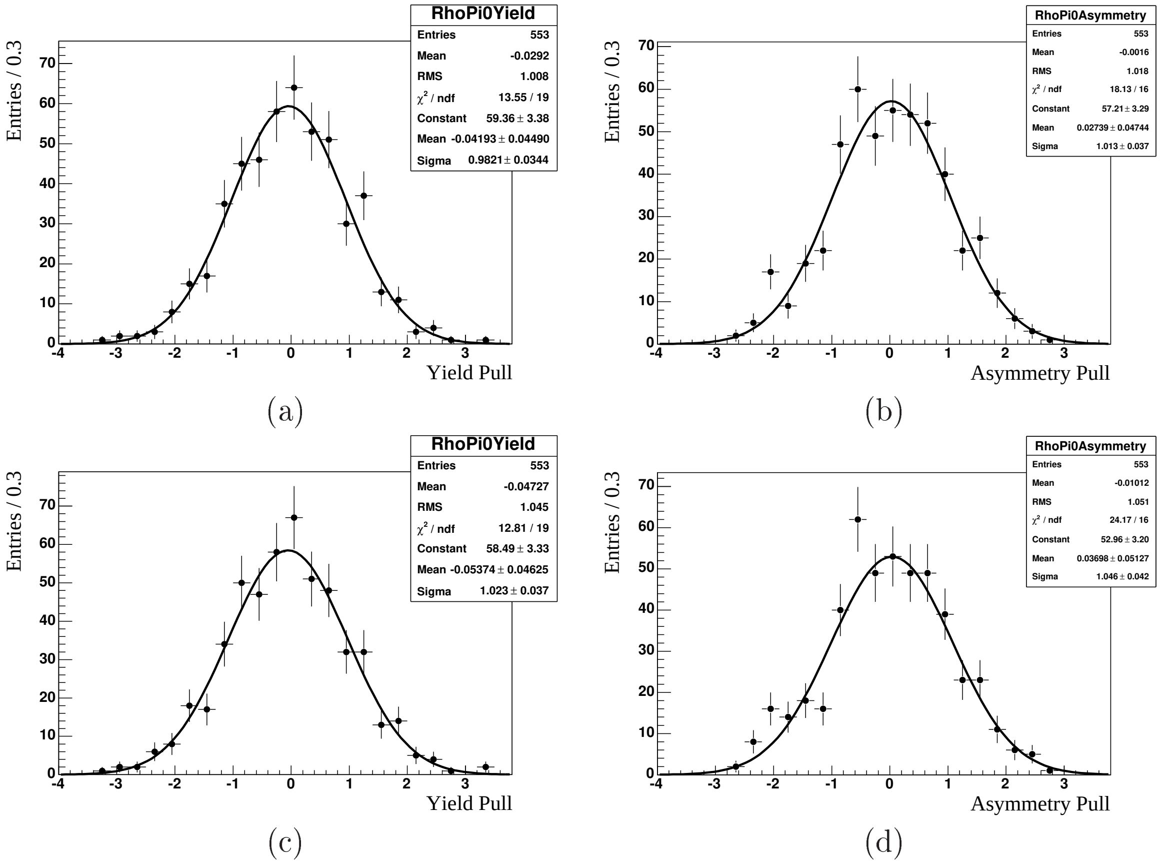


Figure 4.8: $B^{\pm} \to \rho^{\pm}\pi^{0}$ pull distributions from toy MC studies. The plots show the pulls for the total number of signal (left) and $\mathcal{A}_{CP}$ (right), Superimposed are Gaussian fits to the data. (a) and (b) are for $\mathcal{A}_{CP}$ =0, while (c) and (d) are for $\mathcal{A}_{CP}$ =0.5. Pull distributions for the other non–zero $\mathcal{A}_{CP}$ values are also consistent with 0 bias and unit width.

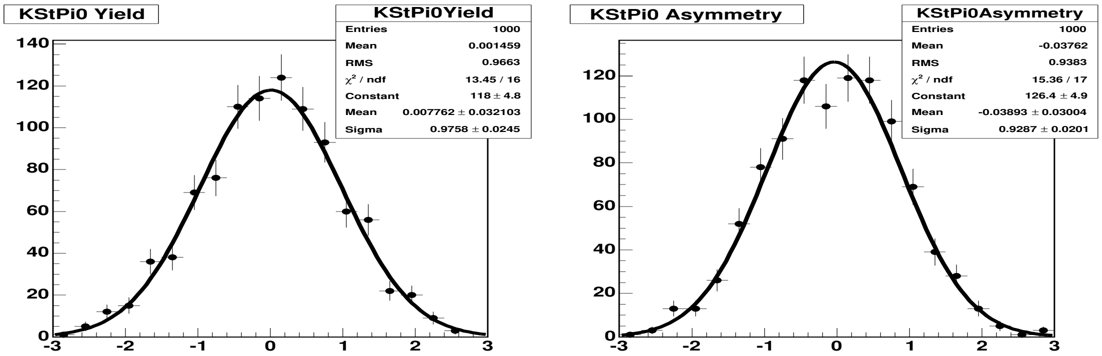


Figure 4.9: Pull distributions from toy MC studies using full Monte Carlo for the signal $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$ mode. The plots show the pulls for the total number of signal (left) and $\mathcal{A}_{CP}$ (right). Superimposed on both are Gaussian fits.

confidence that the fit is functioning as intended, and that the model is fully self–consistent with itself and the MC samples from which the PDFs are formed. It remains to be proven that the model is a true reflection of the real world, and there is one further study that is performed for this final validation before fitting to the data. This study is left until the end of the chapter.

Comparing Figures 4.7 and 4.9, we can see that the error on the asymmetry is overestimated in both cases. The value of the width of the Gaussian fit is about 0.92 in both cases and is significantly different from 1; however no correction is applied for this effect as the error remains on the conservative side and the value of 0.92 is acceptably good.

4.6.4 Optimisation of ANN

It has been shown that the dominant source of background to this analysis comes from light–quark continuum production, and that the most effective way to discriminate against it is with the ANN variable. In order that the highest significance on any observed yields is attained it was decided to optimise both of the analyses using the selection cut–off of the ANN variable.

The optimisation procedure adopted was to maximise the significance of the measured N^{Signal}. To do this a branching ratio of $\mathcal{B}(B^{\pm} \rightarrow K^{*\pm}\pi^{0},\, K^{*} \rightarrow K^{+}\pi^{0})$=$7\times$

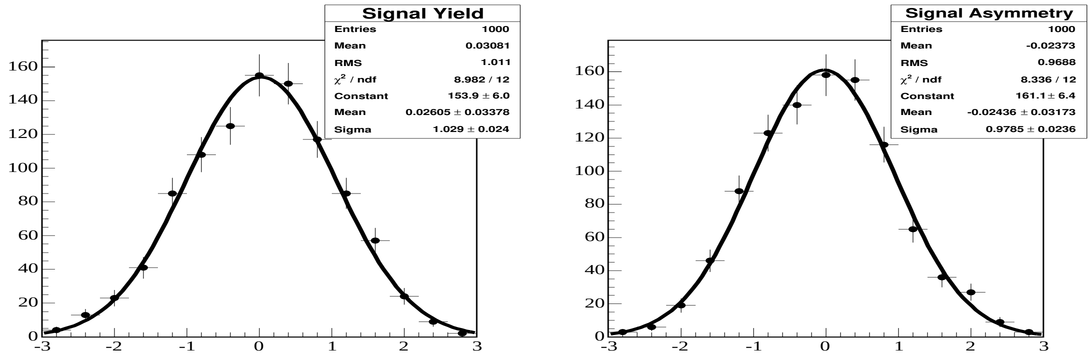


Figure 4.10: Pull distributions from toy MC studies using full Monte Carlo for the signal $B^{\pm} \to \rho^{\pm}\pi^{0}$ mode. The plots show the pulls for the total number of signal (left) and $\mathcal{A}_{CP}$ (right). Superimposed on both are Gaussian fits.

10^{-6} is assumed in the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis, and $\mathcal{B}(B^{\pm} \to \rho^{\pm}\pi^{0})$=$12 \times 10^{-6}$ in the case of the $\rho^{\pm}\pi^{0}$ analysis. Toy studies were performed where the cut–off on the ANN variable was progressively increased to reduce the number of continuum background in the fit. The significance of the measured value of the signal yield was then calculated for each of the cuts on ANN. It was hoped that an optimal cut on ANN can be clearly determined.

In performing the toy studies, the relative efficiencies of the signal and various background components for the various ANN cuts was taken into account by the number of events of each type generated. The ANN variable is varied between 0.2 and 0.9.

The statistical significance is defined in as

$$\text{Significance} = \sqrt{-2\ln(\mathcal{L}_{Null}/\mathcal{L}_{Max})} \tag{4.9}$$

where $\mathcal{L}_{Max}$ is the likelihood value from the fit and $\mathcal{L}_{Null}$ is the likelihood from the fit assuming the null hypothesis, i.e. setting N^{signal} to zero.

For the $B^{\pm} \to K^{*\pm}\pi^{0}$ 1000 samples of toy MC generated data were fitted for each cut on ANN. For each toy study, the significance distribution was fitted with a Gaussian as shown in Figure 4.11. The mean of the significance distribution for each cut is shown in Figure 4.12. It is clear from this plot that the significance drops for values of $ANN > 0.65$. Below this value though, the significance is greatest for the lowest value of ANN considered. In order to maintain high significance a cut of $ANN > 0.25$ was chosen.

Similarly for the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis, 1000 samples of toy MC generated data were fitted for each cut on ANN. In this case, the significance starts to decrease slightly below around 0.4. Consequently, a cut of $ANN > 0.40$ was chosen as the optimum value.

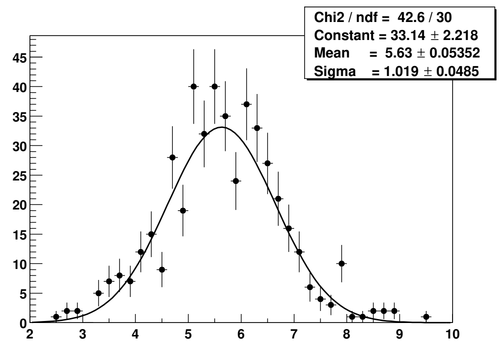
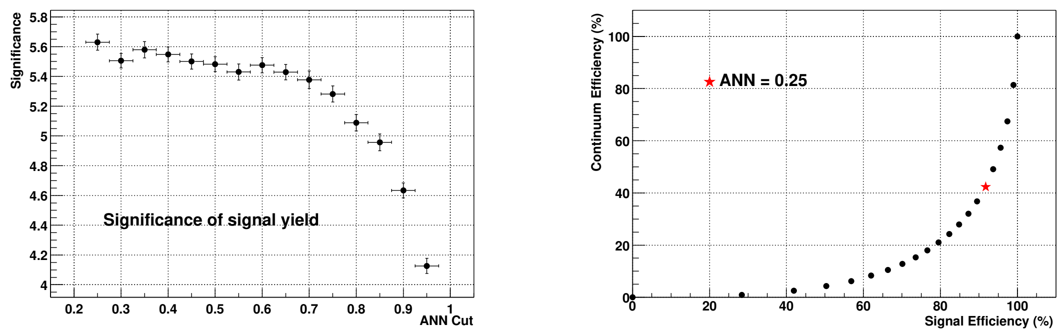


Figure 4.11: A Gaussian fit to the distribution of significances for the $B^{\pm} \to K^{*\pm}\pi^0$ yield assuming $\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^0,\, K^* \to K^+\pi^0)$=$7 \times 10^{-6}$ and with a cut on $ANN = 0.25$.



(a) (b)

Figure 4.12: (a) The mean of the significance distributions for the different cuts on ANN applied for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis. The significance of the cut is reasonably flat for all cuts on ANN below 0.6. It should be noted that the points are correlated. (b) shows the efficiency of the ANN cut applied with respect to all other cuts for continuum versus $K^{*\pm}\pi^0$ signal MC. Each point corresponds to an increasing value of the cut on ANN, with $ANN = 0$ corresponding to 100% efficiency. The star corresponds to the value chosen for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, $ANN > 0.25$. An analogous investigation for the $B^{\pm} \to \rho^{\pm}\pi^0$ yields an optimal cut of $ANN > 0.40$

4.7 Blind fits to data

There is one further validation study of the fit model that is wise to undertake. This is to examine the distribution of likelihoods from toy MC experiments which are seeded with the fitted parameters from a *blind* fit to the data. The purpose of the blind approach is to avoid influencing the direction of the analysis in ways that may compromise its impartiality. It would be unsatisfactory to conclude the model is wrong based on *a posteriori* knowledge or assumptions of what the results 'should' be, founded on some degree of prior expectation. By examining the distribution of the likelihoods, the only conclusion one may draw is that the model is likely or unlikely to represent the data. No further information can be extracted about the inconsistencies, therefore one is not tempted to change the model in a specific way such that it fits the data better.

Mechanically the blind approach to this validation is implemented by caching the results of the fit to the data. The information in the cache is then read by the software used to generate toy MC datasets, but is otherwise kept unreadable. The toy MC datasets are created with the signal and continuum asymmetries and yields fixed to their fitted values, with the usual Poisson smearing applied to the yields. The continuum component of the datasets are generated with the continuum PDF constructed from the fitted parameter values.

The toy datasets are fitted in an exactly analagous way to the blind data fit, with the results also kept hidden. The only output that is stored in each experiment is the value of the minimum log–likelihood obtained from fitting to that particular toy dataset. If the model describes the data well the value of the maximum likelihood from the fit to the data should be close to the peak of this distribution. Figure 4.13 shows the distribution of log–likelihoods from the toy studies with the value of the log–likelihood from the data fit indicated, for both analyses. As can been seen in each case, the value of the log–likelihood from the data fit is close to the mean of the distribution, thereby giving confidence that the model describes the data well.

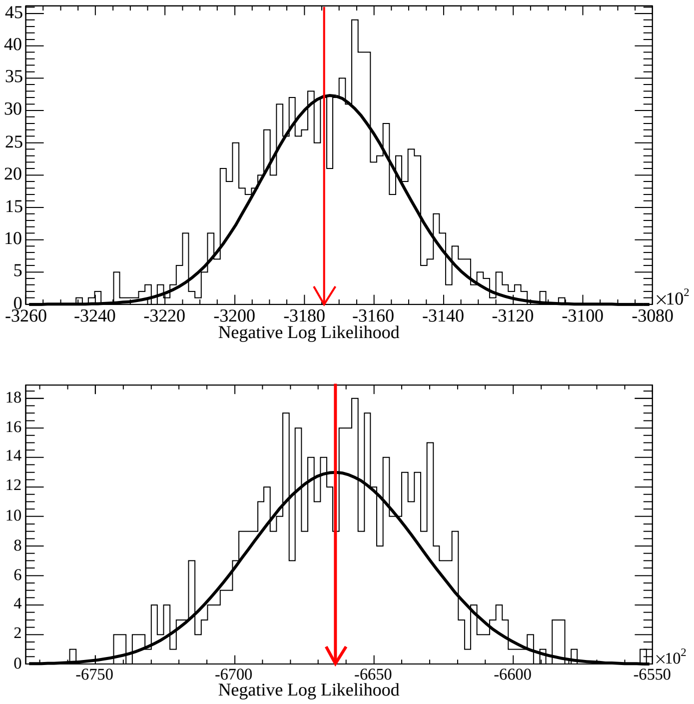


Figure 4.13: Distribution of likelihoods from toy MC experiments seeded with values from a blind fit to the data. The top plot is for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, the bottom is for $B^{\pm} \to \rho^{\pm}\pi^0$. The values of the likelihood from the fit are shown as vertical arrows, indicating the model describes the data well.

Chapter 5

Results and conclusions

5.1 Introduction

This chapter presents the systematic uncertainties which affect the analyses, the final results of the branching fraction and direct CP asymmetry measurements for the decays $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ and $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$, and the conclusions.

5.2 Analysis of systematic uncertainties

The following section details the systematic uncertainties which affect this analysis. It is split into three sections: The first section describes the systematic uncertainty associated with the yields, which affects the significance of the result. The second section describes the systematic uncertainties when calculating the branching fraction from the yields. The third section deals with the systematic uncertainty on the asymmetry measurement.

The systematic uncertainties associated with background estimation and the model are all evaluated by refitting the data.

5.2.1 Systematic uncertainties on the yields

Systematic uncertainties from background estimation

Although the continuum background yield is allowed to float in the fit to the data, the backgrounds from B–related sources identified from simulation are fixed to the levels determined from their efficiencies and assumed branching fractions as shown in Tables 4.4 and 4.5.

Some of the individual B backgrounds have been well measured and for these modes the number of events fixed in the fit is varied by assuming a branching fraction change of $\pm 1\sigma$ where σ is the quoted error on the measurement. In some cases there is only an upper limit available for the decay, and in these instances the branching fraction assumed is half the upper limit. These modes have their branching fractions varied by $\pm 100\%$ of their assumed value (i.e. $\pm 50\%$ of the upper limit). For the generic B component we fix the rate based on the efficiency seen in the generic B Monte Carlo sample, and vary the amount based on the statistical uncertainty of the number. In the $B^{\pm} \to K^{*\pm}\pi^0$ ($B^{\pm} \to \rho^{\pm}\pi^0$) analysis the higher $K^{**\pm}$ ($\rho^{*\pm}$) contributions are included in the fit at the level described in Section 4.5.3. An uncertainty of 100% is assigned to that number. The shifts in the fitted signal yields are found for each mode in turn and then added in quadrature to find the total systematic error. The results of the five biggest background contributors to the the variation in signal yield are shown, for each analysis, in Tables 5.1 and 5.2. In the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, the largest individual contribution to the systematic uncertainty comes from the $K^{**\pm}$ estimation; in the case of the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis it is from the $B \to a_1\pi^0$, not the $\rho^{*\pm}$.

The total effects on the fitted number of signal events are:

$$\Delta N^{Signal}_{Norm}(B^{\pm} \to K^{*\pm}\pi^0) \quad = \quad {}^{+14.1}_{-13.5}$$

$$\Delta N^{Signal}_{Norm}(B^{\pm} \to \rho^{\pm}\pi^0) \quad = \quad {}^{+12.4}_{-11.5}$$

$B^{\pm} \to K^{*\pm}\pi^0$ Analysis		
Background mode	Variation in normalisation	Variation in fitted yield
$B^{\pm} \to K_0^{*\pm}(1430)\pi^0$	± 34	${}^{+13.5}_{-13.0}$
$B^0 \to K^{*\pm}\rho^{\mp}$	± 32.4	${}^{+2.3}_{-2.2}$
$B^{\pm} \to \rho^{\pm}\pi^0$	± 3.2	${}^{+2.0}_{-1.9}$
$B^0 \to \pi^0 K^{*0}$	± 5.8	${}^{+1.5}_{-1.5}$
$B^0 \to K^{\pm}\rho^{\mp}$	± 7.9	${}^{+1.3}_{-1.3}$

Table 5.1: Table of significant contributions from background normalisations to the systematic uncertainty on the $B^{\pm} \to K^{*\pm}\pi^0$ signal yield.

$B^{\pm} \to \rho^{\pm}\pi^0$ Analysis		
Background mode	Variation in normalisation	Variation in fitted yield
$B^0 \to a_1\pi^0$	± 47	${}^{+8.4}_{-10.0}$
$B^0 \to \rho^+\rho^-$	± 30	${}^{+6.0}_{-7.7}$
$B^{\pm} \to \rho^{*\pm}\pi^0$	± 8	${}^{+2.5}_{-3.0}$
$B^0 \to \rho^+\pi^-$	± 14	${}^{+2.5}_{-2.0}$
$B^{\pm} \to \rho^{\pm}\gamma$	± 1.4	${}^{+1.4}_{-0.5}$

Table 5.2: Table of significant contributions from background normalisations to the systematic uncertainty on the $B^{\pm} \to \rho^{\pm}\pi^0$ signal yield.

Systematic uncertainties from modelling

While the continuum PDF parameterisations are determined from the fit to data, the PDF shapes for each of the B background and signal contributions are determined from MC simulation. The limitations with this approach are that one relies entirely on the accuracy of the simulation to determine the distributions of the fit variables. However, the MC samples have already undergone corrections in a variety of different manners for precisely this purpose, although of course one can never guarantee that these corrections necessarily propagate through to the variables considered in the fit. Additionally, there exist no good control decay modes whereby one can isolate a large, high purity sample of known decays to a track and two π^0 mesons, making it difficult to verify the relevant distributions in data. There are thus no explicit corrections adopted in these analyses to the MC PDF shapes to account for this, except in the isolated case of the truth–matched signal ΔE shape, as is this is known to have a definite data:simulation difference in many charmless analyses. This correction is discussed below.

Additionally to this, many modes do not have sufficient statistics in the samples used to ensure that statistical variations – especially in the tails of distributions – do not have an effect in the overall modelling of the PDFs. To attempt to account for this uncertainty, the PDFs are remade in two ways:

- *Parameterised PDFs*: When the PDF has a parametric form, the values of the fixed parameters are varied within the errors of their fitted value from simulation. Care is taken to account for any correlations between parameters by using the parameter error matrix. The eigenvectors of the error matrix represent the directions in the parameter space along which the variation should take place, and the degree of variation is essentially $\pm 1\sigma$ in this direction. The exact details of the error variation are given in [65]. In cases where the error matrix is approximately diagonal, the parameters can be treated independently and varied as such. The two cases where there are fixed–shape parametric distributions are:

  i The signal resonance shapes. These are well described by a Breit–Wigner lineshape with two parameters - the resonance pole and the width.

  ii The Crystal Ball lineshape used to describe the transformed ANN distributions of B backgrounds and signal.

  Some continuum lineshapes have a parametric form, but the parameters are always floated in the fit and therefore have no modelling systematic associated to them.

- *Keys PDFs*: When *Keys* shapes are used, the PDFs are remade with the bandwidth 'smeared' to double the value used for the PDFs in the nominal fit. This amounts to a shift from 0.7 to 1.4 in the two–dimensional case (ΔE, m_{ES}), and 1.0 to 2.0 for the one–dimensional shapes. The effect of this is to smooth the PDF and broaden strong peaks. As was explained in Section 4.3.2, the initial choice of bandwidth is made by an inspection of the resulting shape, so that the bandwidth is not so small as to cause the PDF to display a 'spiking' stucture associated with statistical fluctuations in the MC sample used to create it. It does not therefore make sense to halve the bandwidth for the systematic study as one would certainly be open to modelling these fluctuations. The systematic error obtained from doubling the bandwidth is applied symmetrically in the absence of an obvious strategy equivalent to a symmetric variation of the bandwidth

The PDF descriptions for each variable are independent, as are the PDFs for individual modes. Thus the fit to the data has to be rerun once for every variation of every parametric PDF, every 1–D KEYS PDF and every 2–D KEYS PDF. This amounts to rerunning the fit over 50 times in both analyses. The changes in the nominal yield are added in quadrature to give the final uncertainty due to modelling. The results are summarised below:

$$\Delta N^{Signal}_{Model}(B^{\pm} \to K^{*\pm}\pi^{0}) = {}^{+1.5}_{-3.9}$$

$$\Delta N^{Signal}_{Model}(B^{\pm} \to \rho^{\pm}\pi^{0}) = {}^{+4.7}_{-4.2}$$

Systematic uncertainties from ΔE

It is possible for the signal peak in ΔE to be different in the data compared to its position in the MC sample. The two-body $B^{\pm} \to \pi^{\pm}\pi^{0}$ analysis at *BABAR* [66] records a systematic shift of 5 MeV in the peak position of ΔE between data and MC simulation [67]. This analysis is unable to undertake a similar study, for the statistics in the final data samples do not allow a fit of the mean position of the signal ΔE component. To estimate the effect of such a shift, the PDFs over the ΔE distribution for signal are shifted by $\pm$ 5 MeV and the data is refitted. The change in yields for the two analyses are summarised below:

$$\Delta N^{Signal}_{\Delta E}(B^{\pm} \to K^{*\pm}\pi^{0}) = \pm 2.2$$
$$\Delta N^{Signal}_{\Delta E}(B^{\pm} \to \rho^{\pm}\pi^{0}) = \pm 2.6$$

Systematic uncertainties from SCF and mistag fractions

The analyses use the fraction of SCF and, in the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis the fraction of mistag, as fixed parameters within the fit as determined from the MC simulation. There are a number of effects that can contribute to these numbers being wrong. Firstly the truth–matching algorithm is unlikely to be 100% efficient. However this is not so problematic, for properly reconstructed candidates that get assigned to the self–cross–feed component will be taken up implicitly in the PDF description. The main source of uncertainty then comes from the best candidate selection and the relative probability of selecting a properly reconstructed candidate. This obviously has some basis on the difference between the data and MC simulation multiplicity distributions for the signal final states which are heavily influenced by the distributions of low energy, background π^{0} mesons. Studies by the neutrals working group [68] on data:MC differences for the distributions of low energy π^{0} mesons indicate that a conservative uncertainty to assign to the SCF fraction is $\pm 10\%$. Similarly for the mistag, the main source of uncertainty comes from the MC simulation of low energy charged π mesons and the efficiency of their detection. Although the data:MC simulation agreement is better for the tracking [69] than for the neutrals, a similarly conservative error of $\pm 10\%$ is assigned to the mistag fraction.

The shifts in the fitted number of signal events due to these effects were found to be

$$\Delta N^{Signal}_{SCF}(B^{\pm} \to K^{*\pm}\pi^{0}) = \pm 1.8$$

$$\Delta N^{Signal}_{SCF}(B^{\pm} \to \rho^{\pm}\pi^{0}) = \pm 12.2$$

$$\Delta N^{Signal}_{Mistag}(B^{\pm} \to \rho^{\pm}\pi^{0}) = \pm 2.0$$

Systematic corrections from fit procedure

As was described in Chapter 4, toy studies indicated that the fit was unbiased – the pull distributions had means consistent with zero. Therefore no systematic correction is applied to the central values of the fitted yields or asymmetries.

Summary of systematic uncertainties on the yields

A summary of the systematic uncertainties on the yields is shown in Table 5.3

5.2.2 Systematic uncertainties on the branching fraction determination

Systematic uncertainties on the calculation of the branching fractions $\mathcal{B}$ are listed in Table 5.4 and detailed below. Corrections are applied to the MC simulation such that there is better agreement with the data. These corrections inevitably have an associated error. This section discusses the sources of the errors and the effect they have on the efficiency. The individual contributions are detailed below. Finally, the error due to the original number of B mesons in the sample is given.

Neutral efficiency corrections

The recipe used to correct the π^0 energy distribution has a systematic error associated with it, and is provided by the Neutrals Working Group [68]. The recommendation is for an uncertainty of $\pm 3\%$ on the efficiency be added linearly for

Absolute Systematic Errors on Yields ($B^\pm \to K^{*\pm}\pi^0$)	
Source	$\sigma^{Yield}_{Syst.}(B^\pm \to K^{*\pm}\pi^0)$ (Events)
Background Normalisation	$^{+14.1}_{-13.5}$
PDF Shapes	$^{+1.5}_{-3.9}$
ΔE shift	± 2.2
SCF fraction	± 1.8
Total	$^{+14.5}_{-14.3}$
Absolute Systematic Errors on Yields ($B^\pm \to \rho^\pm\pi^0$)	
	$\sigma^{Yield}_{Syst.}(B^\pm \to \rho^\pm\pi^0)$ (Events)
Background Normalisation	$^{+12.4}_{-11.5}$
PDF Shapes	$^{+4.7}_{-4.2}$
ΔE shift	± 2.6
SCF fraction	± 12.2
Mistag fraction	± 2.0
Total	$^{+18.3}_{-17.6}$

Table 5.3: Table of sources of systematic errors on the yields. The total systematic error on the yields is the sum of the individual contributions added in quadrature.

each π^0 with a $\pm 2\%$ error per π^0 to be added in quadrature to account for the variation in π^0 efficiency over the momentum range. Adding these in quadrature ($\sqrt{[3+3]^2 + [2+2]^2}$)gives the total relative systematic error due to the π^0 efficiency differences to be 7.2%.

Systematics from track PID

There is no standard procedure for estimating systematic uncertainties due to PID. The PID corrections are derived from control samples of fully reconstructed decays such as $D^{*+} \to \pi^+ D^0 (\to K^-\pi^+)$. This decay for instance yields a sample of negative kaons and a sample of positive pions over some momentum distributions appropriate to this event topology. By applying the PID selection requirements to these samples one can determine the efficiency of the chosen selector for the given species, at a given momentum, in a given region of the detector acceptance, *etc.* Comparisons between the efficiency of PID in the MC simulation and the efficiency of PID in the data are used to build the PID tables prescribing the efficiency correction probabilities for a given selector, species, charge, and detector region. Using these tables incurs a systematic error due to the uncertainty in the probabilities from the statistics used to calculate them. However, these tables are only really 'true' for the decay mode used to make them and in general the data:MC differences will be different for differing event topologies [70]. The 'correct' procedure to estimate the systematics due to PID is to perform a fairly involved study comparing the performance of the chosen PID selector on the given analysis using MC truth, and relating the efficiencies obtained therein to the efficiencies in the PID tables. To the degree that the MC efficiency for PID on the given analysis obtained from truth agrees with that for the MC control sample, it can be reasonably assumed that the analysis procedure applied to real data has a similar PID efficiency to the real data control sample. However, in the absence of this study (which is only necessary if knowing PID systematics accurately is critical to the analysis), the recommendation of the PID Working Group [71] to allow for the disparity between PID efficiencies between the control sample and a 'typical' analysis is put at 1%. As this analysis is dominated by the neutrals correction at 7.2%, the value of 1% for the PID uncertainty is adopted here.

The additional systematic to be applied is to integrate the PID correction uncertainties from the table over the momentum and acceptance of the final sample of tracks. This is estimated to be 1% in both analyses. To be conservative these two systematics are added linearly rather than in quadrature to give a total error of 2% from PID. The same value is used for both analyses.

Systematics from tracking efficiency

The Tracking Efficiency Task Force has detailed recommendations on what systematic error to assign to tracks from the `GoodTracksLoose` list, ranging from 0.8% to 3.5% depending on event multiplicity and whether the correction tables were used [69]. The systematic error on the efficiency for a single charged track when the tracking correction tables are used is 0.8%, and is the error used here.

Total systematic error due to efficiency calculation

Adding all the above contributions in quadrature gives a total relative systematic error on $\mathcal{B}$ to be:

$$\sigma^{\mathcal{B}}_{\mathrm{Eff}} = \pm 7.5\%$$

where the dominant effect comes from the uncertainty in the efficiency of the π^0 reconstruction. This uncertainty is applied for both analyses.

Systematics from B counting

It is assumed that the branching fraction for $\Upsilon(4S) \rightarrow B^+B^-$ is the same as for $\Upsilon(4S) \rightarrow B^0\overline{B}^0$. The error on the total number of B events from B counting is 1.1% [72]. This is taken as the relative error on $\mathcal{B}$ due to this effect.

Summary of relative systematics uncertainties affecting $\mathcal{B}$

Table 5.4 summarises the relative systematics uncertainties affecting the calculation of the branching fraction $\mathcal{B}$.

5.2.3 Systematic uncertainties on $\mathcal{A}_{CP}$

Systematic effects from detector asymmetry

The cross section for the interaction of kaons with protons and neutrons differs with charge. K^- generally have a higher interaction cross–section with nuclei compared to K^+ in the momentum range accessible to *BABAR*. This cross section is also momentum–dependent with the K^-p cross section rising to more than ~ 6 times

Relative Systematic Errors on $\mathcal{B}$	
Source	$\sigma^{\mathcal{B}}_{Syst.}$ ($B^{\pm} \to K^{*\pm}\pi^{0}$ and $B^{\pm} \to \rho^{\pm}\pi^{0}$)
Neutral efficiency	$\pm 7.2\%$
PID Efficiency	$\pm 2\%$
Tracking Efficiency	$\pm 0.8\%$
Total Efficiency	$\pm 7.5\%$
B Counting	± 1.1
Total Relative Systematic on $\mathcal{B}$	$\pm 7.6\%$

Table 5.4: Table of sources of relative systematic errors on the branching fraction. The total systematic error on the branching fraction is the sum of the individual contributions added in quadrature.

the value of the $K^{+}p$ cross section [15]. The source of the asymmetry is due to the presence of the $\overline{u}$ quark in the K^{-}, which can annihilate with the u quarks in the nucleon, unlike for the K^{+}.

The detector absorption asymmetry from inelastic collisions is conveniently already built into the MC simulation, and therefore the resulting asymmetry of the final signal $B^{\pm} \to K^{*\pm}\pi^{0}$ MC sample is taken as the systematic shift. The final signal MC sample has 102,373 positively charged candidates and 99,149 negatively charged candidates. The statistical error on these numbers is about 310 from counting statistics, thus the systematic shift due to the detector asymmetry is significant and estimated to be $(99 - 102)/201 = -0.015$. This asymmetry shift should be seen in the data, so 0.015 needs to be added to the final value of $\mathcal{A}_{CP}$ for $B^{\pm} \to K^{*\pm}\pi^{0}$. The error on this asymmetry correction was considered to be negligible, since a poor simulation of this effect is largely taken up by the PID correction explained in Section 3.4.2.

Pions do not suffer this asymmetry to the same extent for the isospin (i.e. proton–neutron) asymmetry in the detector is small. The MC simulation supports this assumption with a final signal sample asymmetry of less than -0.005. Therefore the potential systematic shift from detector asymmetry effects is considered in the $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$ analysis only.

Systematic uncertainties from background asymmetries

The asymmetries of the background contributions were shown in Tables 4.4 and 4.5. They corresponded to actual measurements, if they existed. In the cases where no measurements existed, zero asymmetry was assumed. In order to calculate the effects of systematic shifts in these asymmetries the following procedure was adopted. Each mode was varied by $\pm 1\sigma$ where σ is the measured error, and the subsequent shift in the fitted asymmetry was found. For contributions with no measurement, an assumed value of σ of 0.2 was used. The individual shifts were then added in quadrature to find the total systematic error. In the $B^{\pm} \rightarrow K^{*\pm}\pi^{0}$ analysis, the greatest individual contribution comes from the $K^{**\pm}$ estimates.

$$\begin{aligned} \sigma^{\mathcal{A}_{CP}}_{Back}(B^{\pm} \rightarrow K^{*\pm}\pi^{0}) &= \pm 0.040 \\ \sigma^{\mathcal{A}_{CP}}_{Back}(B^{\pm} \rightarrow \rho^{\pm}\pi^{0}) &= \pm 0.014 \end{aligned}$$

Systematic uncertainties from background yields

A small shift was observed in the fitted value of $\mathcal{A}_{CP}$ due to the normalisations of the backgrounds, as supposed to their assumed charge asymmetry. The shifts were found to be

$$\begin{aligned} \Delta A^{Signal}_{Norm}(B^{\pm} \rightarrow K^{*\pm}\pi^{0}) &= \pm 0.008 \\ \Delta A^{Signal}_{Norm}(B^{\pm} \rightarrow \rho^{\pm}\pi^{0}) &= \pm 0.006 \end{aligned}$$

which is incorporated into the systematic uncertainty on the final value of $\mathcal{A}_{CP}$. The smaller effect on the asymmetry uncertainty in the $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ is because of the higher signal statistics available in this mode.

Systematic uncertainties from modelling

A small variation was found in the measured value of $\mathcal{A}_{CP}$ due to the PDF modelling. This was found to be

$$\begin{aligned} \Delta A^{Signal}_{Model}(B^{\pm} \to K^{*\pm}\pi^{0}) &= \pm 0.020 \\ \Delta A^{Signal}_{Model}(B^{\pm} \to \rho^{\pm}\pi^{0}) &= \pm 0.011 \end{aligned}$$

Systematic uncertainties from SCF and mistag fractions

As for the yields, varying the SCF and mistag fractions by $\pm 10\%$ has a potential effect on the fitted asymmetries. These are summarised below

$$\sigma^{SCF}_{\mathcal{A}_{CP}}(B^{\pm} \to K^{*\pm}\pi^{0}) < \pm 0.001$$

$$\begin{aligned} \sigma^{SCF}_{\mathcal{A}_{CP}}(B^{\pm} \to \rho^{\pm}\pi^{0}) &< \pm 0.001 \\ \sigma^{Mistag}_{\mathcal{A}_{CP}}(B^{\pm} \to \rho^{\pm}\pi^{0}) &< \pm 0.001 \end{aligned}$$

The mistag fraction has the (small) effect of diluting the asymmetry measurement, but if the asymmetry measurement is small then the effect is very hard to see. As all these variations in the measured asymmetry are tiny, they are regarded as being negligible.

Summary of systematics uncertainties on $\mathcal{A}_{CP}$

Table 5.5 summarises the systematic uncertainties on the charge asymmetry measurements.

Systematic Errors on Asymmetries	
Source	$\sigma^{\mathcal{A}_{CP}}_{Syst.}(B^{\pm} \to K^{*\pm}\pi^{0})$
Background Normalization	± 0.008
Background Asymmetry	± 0.040
Modelling	± 0.020
Total	± 0.05
	$\sigma^{\mathcal{A}_{CP}}_{Syst.}(B^{\pm} \to \rho^{\pm}\pi^{0})$
Background Normalization	± 0.006
Background Asymmetry	± 0.014
Modelling	± 0.011
Total	± 0.02

Table 5.5: Table of systematic sources of errors and their contributions to the asymmetry. The total systematic error is the sum of the individual contributions added in quadrature.

5.3 Results from fit

This section contains the results of the fit to the full Run 1–4 data set, corresponding to $(231.6 \pm 2.6) \times 10^{6}$ $B\overline{B}$ events.

5.3.1 Fit results for $B^{\pm} \to K^{*\pm}\pi^{0}$

A total of 23465 events were fitted with 11955 positively charged. The correlation matrix calculated by MINUIT for the fitted parameters is shown below:

```
PARAMETER  CORRELATION COEFFICIENTS
NO.  GLOBAL      1      2      3      4      5      6      7
  1  0.17107   1.000 -0.128 -0.115  0.019  0.003 -0.014 -0.028
  2  0.24763  -0.128  1.000  0.013 -0.119  0.002  0.058  0.176
  3  0.11510  -0.115  0.013  1.000 -0.001 -0.001  0.002  0.002
  4  0.12083   0.019 -0.119 -0.001  1.000 -0.005 -0.005 -0.038
  5  0.26817   0.003  0.002 -0.001 -0.005  1.000 -0.268 -0.001
  6  0.27387  -0.014  0.058  0.002 -0.005 -0.268  1.000  0.010
  7  0.17655  -0.028  0.176  0.002 -0.038 -0.001  0.010  1.000
```

where the parameters numbered 1–7 correspond to the following:

1. $\mathcal{A}_{CP}^{Signal}$,
2. N^{Signal},
3. $\mathcal{A}_{CP}^{Cont}$,
4. N^{Cont},
5. Coefficient of first power in ΔE polynomial continuum PDF.
6. Coefficient of second power in ΔE polynomial continuum PDF.
7. Argus parameter shape, ζ, for continuum PDF.

It can be seen that correlations between signal and background modes are generally small, with the largest being between parameters 7 (ξ) and 2 (N^{Signal}) taking a value of 0.176.

The number of fitted signal events was found to be 87 ± 26, where the error is statistical only. This can be translated into a branching fraction using Eq.(5.1):

$$\mathcal{B} \; = \frac{N^{Signal}}{\epsilon N^{B^+B^-}} \tag{5.1}$$

With a signal efficiency $\epsilon = 16.48\%$, this gives the fitted branching fraction to be

$$\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^0, K^{*\pm} \to K^{\pm}\pi^0) = [2.3 \;\; \pm 0.7] \times 10^{-6}$$

where the error is again only statistical.

The fitted asymmetry was found to be:

$$\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^0, K^{*\pm} \to K^{\pm}\pi^0) = 0.02 \;\; \pm 0.29$$

where the error is statistical. The central value must be corrected from the asymmetry expected for the K^+ nuclear interactions of $A^{Kp} = -0.015$ (Section 5.2.3). The corrected asymmetry, including the systematic shift is therefore

$$\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^0, K^{*\pm} \to K^{\pm}\pi^0) = 0.03 \;\; \pm 0.29(Stat.)$$

The parameters for the background characteristation PDFs were found to be in agreement with the initial (expected) values. A summary of the fit results for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis is shown in Table 5.6.

Parameter	Fitted Value
$K^{*\pm}\pi^0$ yield	87 ± 26
$K^{*\pm}\pi^0$ asymmetry	0.02 ± 0.29
Continuum yield	22744 ± 154
Continuum asymmetry	-0.02 ± 0.01
Argus shape	-19.9 ± 0.7

Table 5.6: Summary of fit results for $B^{\pm} \to K^{*\pm}\pi^0$, including the fitted value of the *ARGUS* parameter.

The final results for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis inclusive of systematic uncertainties are shown in Section 5.4.

5.3.2 Fit results for $B^{\pm} \to \rho^{\pm}\pi^0$

A total of 47608 events were fitted with 23789 positively charged. The correlation matrix output from MINUIT for the fitted parameters is shown below:

PARAMETER CORRELATION COEFFICIENTS

NO.	GLOBAL	1	2	3	4	5	6	7	8	9	10	11
1	0.15294	1.000	-0.054	-0.143	0.011	-0.006	0.002	0.002	-0.001	-0.006	-0.015	-0.009
2	0.32862	-0.054	1.000	0.004	-0.175	0.078	-0.027	-0.021	0.008	0.074	0.240	0.115
3	0.14344	-0.143	0.004	1.000	-0.001	0.000	-0.001	-0.001	0.000	0.001	0.001	0.001
4	0.17689	0.011	-0.175	-0.001	1.000	-0.021	0.007	0.006	-0.010	-0.012	-0.061	-0.033
5	0.74996	-0.006	0.078	0.000	-0.021	1.000	0.076	0.049	0.003	-0.001	0.036	0.749
6	0.80897	0.002	-0.027	-0.001	0.007	0.076	1.000	-0.795	0.001	0.004	-0.015	0.089
7	0.80771	0.002	-0.021	-0.001	0.006	0.049	-0.795	1.000	-0.001	0.003	-0.014	0.053
8	0.13779	-0.001	0.008	0.000	-0.010	0.003	0.001	-0.001	1.000	-0.136	0.008	0.005
9	0.15655	-0.006	0.074	0.001	-0.012	-0.001	0.004	0.003	-0.136	1.000	0.020	0.003
10	0.24394	-0.015	0.240	0.001	-0.061	0.036	-0.015	-0.014	0.008	0.020	1.000	0.047
11	0.75631	-0.009	0.115	0.001	-0.033	0.749	0.089	0.053	0.005	0.003	0.047	1.000

where the parameters numbered 1–11 correspond to the following:

1. $\mathcal{A}_{CP}^{Signal}$,
2. N^{Signal},
3. $\mathcal{A}_{CP}^{Cont}$,

4. N^{Cont},

5. Value of exponent in ANN_{Tran} exponential PDF.

6. Coefficient of first power in ANN_{Tran} polynomial continuum PDF.

7. Coefficient of second power in ANN_{Tran} polynomial continuum PDF.

8. Coefficient of first power in ΔE polynomial continuum PDF.

9. Coefficient of second power in ΔE polynomial continuum PDF.

10. Argus parameter shape, ζ, for continuum PDF.

11. The ratio R of the relative normalisations of the exponential:polynomial components in the continuum PDF for ANN_{Tran} (see Section 4.3).

The number of fitted signal events was found to be 367 ± 50, which converts to a branching fraction of:

$$\mathcal{B}(B^{\pm} \to \rho^{\pm}\pi^{0}) = [10.3 \ \pm 1.4] \times 10^{-6}$$

for an efficiency of 15.4%. The quoted error is statistical only. As for the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis, the parameters for the background characteristation PDFs were found to be in agreement with the expected values from simulation. A summary of the fit results for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis is shown in Table 5.7.

The fitted asymmetry was found to be:

$$\mathcal{A}_{CP}(B^{\pm} \to \rho^{\pm}\pi^{0}) = 0.03 \ \pm 0.13$$

where the error is statistical. Unlike for the $B^{\pm} \to K^{*\pm}\pi^{0}$ case there is no systematic shift correction to the asymmetry from detector effects.

Parameter	Fitted Value
$\rho^{\pm}\pi^0$ yield	367 ± 50
$\rho^{\pm}\pi^0$ asymmetry	0.03 ± 0.13
Continuum yield	46314 ± 221
Continuum asymmetry	-0.00 ± 0.01
Argus shape	-15.9 ± 0.5

Table 5.7: Summary of fit results for $B^{\pm} \to \rho^{\pm}\pi^0$, including the fitted value of the *ARGUS* parameter.

5.3.3 Stability cross-check

As a stability check, the fits were redone without the mass variable used in the fit. If the signal yield were to change subtantially, in particular if it significantly increased, this would be suggestive of some significant background unaccounted for which shares the same or similar final state but is not peaked in the $h^{\pm}\pi^0$ invariant mass. The results of the fit without the mass included for both analyses are

$$\text{Yield}(B^{\pm} \to K^{*\pm}\pi^0) = 88 \pm 29$$

$$\text{Yield}(B^{\pm} \to \rho^{\pm}\pi^0) = 352 \pm 51$$

These results are consistent with the full fit using the mass, giving confidence there are no significant missing contributions.

5.4 Final results inclusive of systematics

For the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, the absolute systematic error on the yield is, from Table 5.3, $^{+14.5}_{-14.3}$ events. This translates to a relative systematic error of $^{+16.7}_{-16.4}\%$. Similarly for the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis, the absolute systematic error on the yield is $^{+18.3}_{-17.6}$ events, translating to a relative systematic error of $^{+5.0}_{-4.8}\%$. The total relative systematic error on the branching fraction is thus the quadrature sum of the relative systematic error on the yield, and the relative systematic error due to the efficiency and B counting, as given in Table 5.4.

The significance of the branching fraction, not including the systematic errors is determined from the following equation:

$$\text{Significance} \;=\; \sqrt{-2\log(\mathcal{L}_{Null}/\mathcal{L}_{Max})}$$

The values of the likelihoods from the final fit and the fit assuming the null hypothesis for signal gives a statistical significance of:

$$\text{Significance} \;=\; 4.2\sigma$$

Including systematic uncertainties[1] the significance of the result is

$$\text{Significance including systematics} \;=\; 3.8\sigma.$$

The final fit results for the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis including systematic errors are given below.

$$\begin{aligned}
\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^{0}, K^{*\pm} \to K^{\pm}\pi^{0}) &= [2.3\ \pm 0.7\ (Stat.)\ \pm 0.4\ (Syst.)] \times 10^{-6} \\
\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^{0}) &= 0.03\ \pm 0.29\ (Stat.)\ \pm 0.05\ (Syst.)
\end{aligned}$$

From isospin symmetry one may write the K^{*} state as

$$\left|\frac{1}{2},\frac{1}{2}\right\rangle = \sqrt{\frac{1}{3}}\left(\left|\frac{1}{2},\frac{1}{2}\right\rangle + |1,0\rangle\right) + \sqrt{\frac{2}{3}}\left(\left|\frac{1}{2},-\frac{1}{2}\right\rangle + |1,1\rangle\right)$$

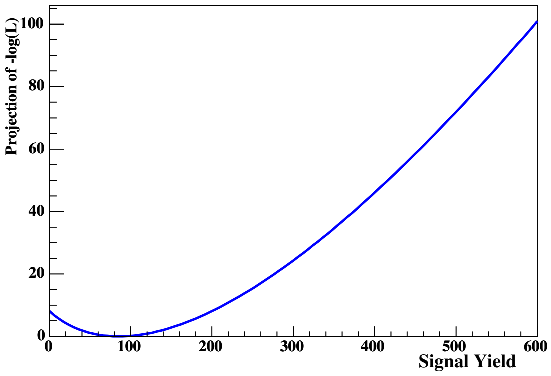
where the first term on the right hand side represents the isospin decomposition of the $K^{+}\pi^{0}$ state and the second term that of the $K^{0}\pi^{\pm}$. The assumption that the branching fraction of $(K^{*\pm} \to K^{\pm}\pi^{0})$ is 1/3 implies the total branching fraction is

$$\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^{0}) = [6.9 \pm 2.1\ (Stat.) \pm 1.2\ (Syst.)] \times 10^{-6}.$$

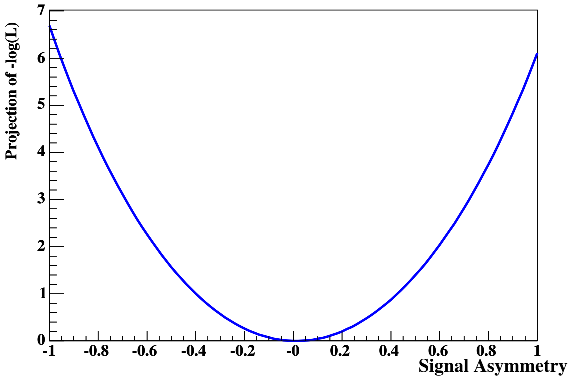
The final fit results for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis including systematic errors are given below.

$$\begin{aligned}
\mathcal{B}(B^{\pm} \to \rho^{\pm}\pi^{0}) &= [10.3\ \pm 1.4\ (Stat.)\ \pm 0.9\ (Syst.)]\ \times 10^{-6} \\
\mathcal{A}_{CP}(B^{\pm} \to \rho^{\pm}\pi^{0}) &= 0.03\ \pm 0.13\ (Stat.)\ \pm 0.02\ (Syst.)
\end{aligned}$$

[1]The value of the signal yield is fixed to the lower bound from the systematic uncertainties associated with the yield only, and the fit is redone.



(a)



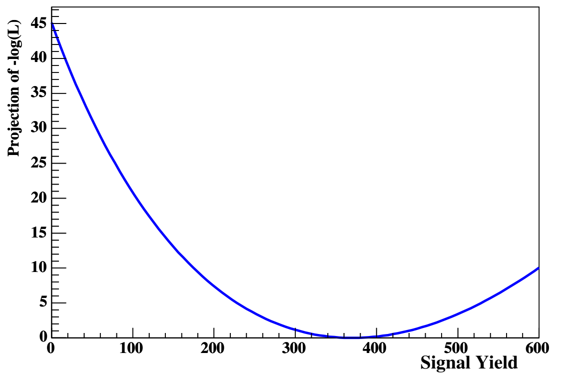
(b)

Figure 5.1: Likelihood curves for (a) N^{Signal} and (b) A^{Signal} for the $B^{\pm} \to K^{*\pm}\pi^{0}$ analysis.

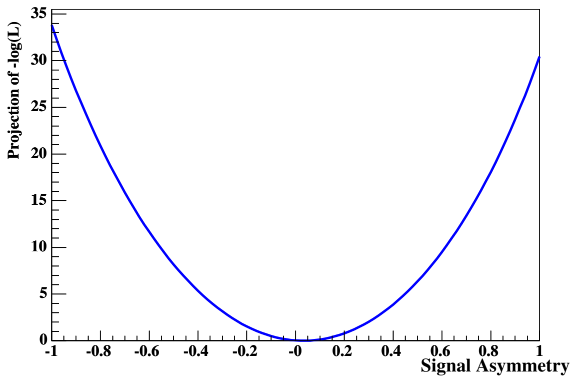
The significance of the $B^{\pm} \to \rho^{\pm}\pi^{0}$ result including systematic effects is 8.7. Figures 5.1 and 5.2 (a) and (b) show the likelihood curves for each of the fitted number of signal and asymmetry respectively for the two respective analyses.

5.4.1 Signal–enhanced discriminating variable projections

To give confidence in the results, it is useful to perform a visual check of the distributions in real data of the discriminating variables with the appropriately normalised component PDFs overlaid. If the likelihood fit yields N signal events, there should appear to be approximately N signal-like events peaking above the backgrounds in all the variables considered. However, given that these analyses are both dominated by large continuum backgrounds, any projection using the full data set will not show significant signal that can be discerned by eye. Instead, the data set is reduced in



(a)



(b)

Figure 5.2: Likelihood curves for (a) N^{Signal} and (b) A^{Signal} for the $B^{\pm} \rightarrow \rho^{\pm}\pi^{0}$ analysis.

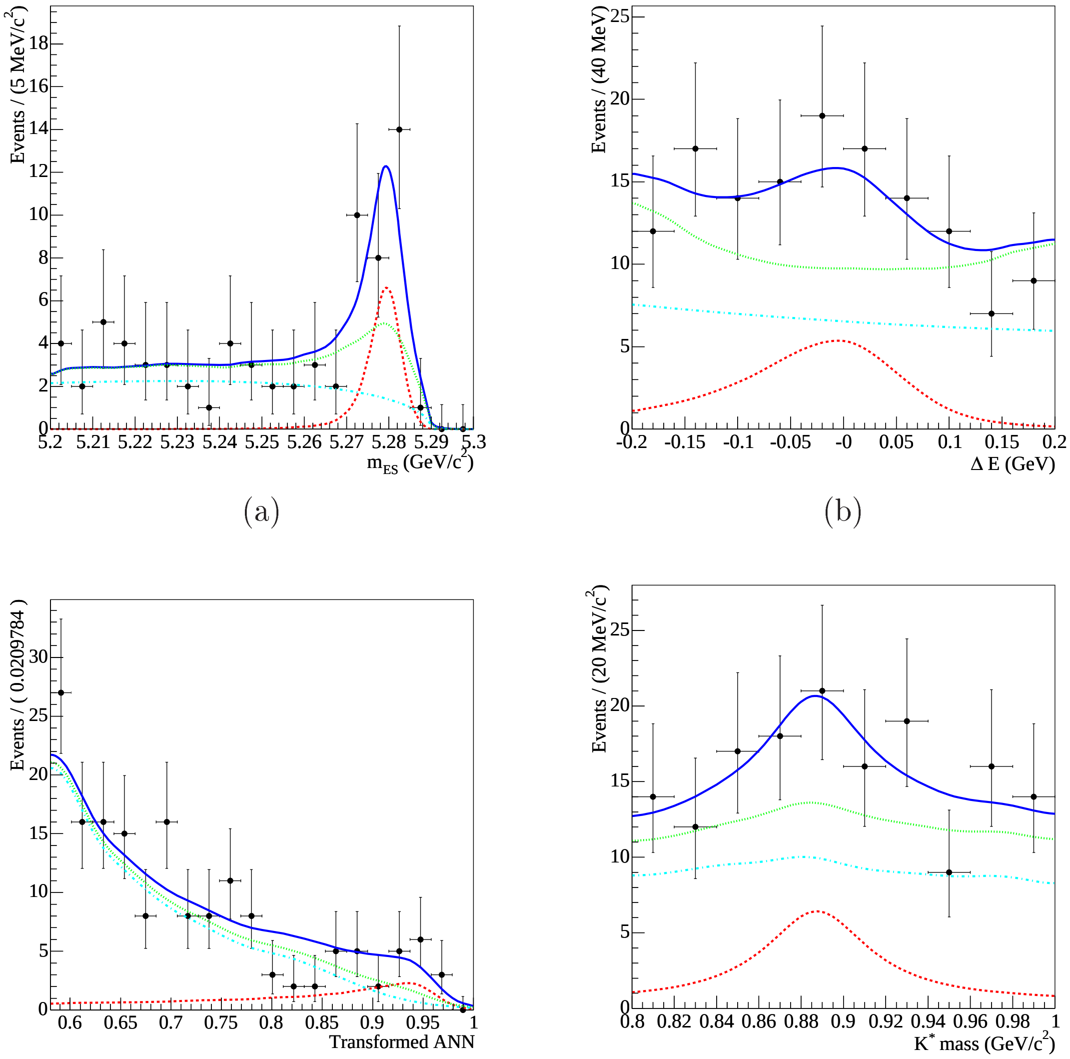
size in such a way that signal events appear enhanced when making the plots. This is done by making an event by event cut on the likelihood ratio R, defined as

$$R(e) = \frac{N_{signal}\mathcal{P}_{signal}(e)}{N_{signal}\mathcal{P}_{signal}(e) + \sum_i N_i^{BG}\mathcal{P}_i^{BG}(e)} \tag{5.2}$$

where e is the event index, N_i are the central values of the yields from the fit and $\mathcal{P}_i$ are the PDFs with the projected variable *integrated out*. For example, if the plot was to be of m_{ES}, no information about m_{ES} would be used to make the plot (otherwise one would simply be selecting events which fall under the peak in m_{ES}). What the likelihood cut does is effectively select events which fall under the peaks in the other three variables (i.e. events with a high value of the likelihood), and if all is well with the fit these events should fall under the peak in m_{ES} too. In making the cut, there is an associated efficiency loss suffered for each of the background modes and the signal. This efficiency, ϵ_{mode}, is determined from MC simulations, and the overlaid PDFs are normalised in the plots by the value $N_{\text{mode}}.\epsilon_{\text{mode}}$. The value of the cut chosen for each plot is not optimised, but is based on judgement.

The projections from the $B^{\pm} \to K^{*\pm}\pi^0$ analysis of m_{ES}, ΔE, the transformed ANN variable and the K^* mass are shown in Figure 5.3 (a-d), whilst the equivalent projections for the $B^{\pm} \to \rho^{\pm}\pi^0$ analysis are shown in in Figure 5.4(a-d).

The error bars in the plot are asymmetric due to the nature of the Poisson distribution and are not simply $\sqrt{N}$ errors. To obtain these error bars one integrates the Poisson distribution about the mean until one gets 34% on each side (representing 68% of the total, and hence 1 σ each side) thereby ensuring that there is no probability overflow below zero. The distribution of points around the fit mean, especially for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis, might be suggestive that these errors are overestimated (the correlation between the points and the fit appears in some cases to be too strong). However, in the case of both m_{ES} and ΔE, the variables do have correlations and integrating the 2-D PDF over the other dimension does not yield a completely independent PDF. It is not unreasonable to suppose an effect of this will be that some events are preferentially selected over another and will therefore



(a) (b) (c) (d)

Figure 5.3: The distributions for m_{ES} (a), ΔE (b), transfromed ANN (c), and $K^{*\pm}$ mass (d) are shown, for candidates passing a cut on the likelihood ratio R (see text). The plots are for the $B^{\pm} \to K^{*\pm}\pi^0$ analysis. These plots do not represent the full information used in the maximum likelihood fit but rather a subset of the data. In each plot the solid line represents the total PDF, the dotted-dashed line represents the continuum contribution, the dotted line represents the continuum + B background, and the dashed line represents the signal component.

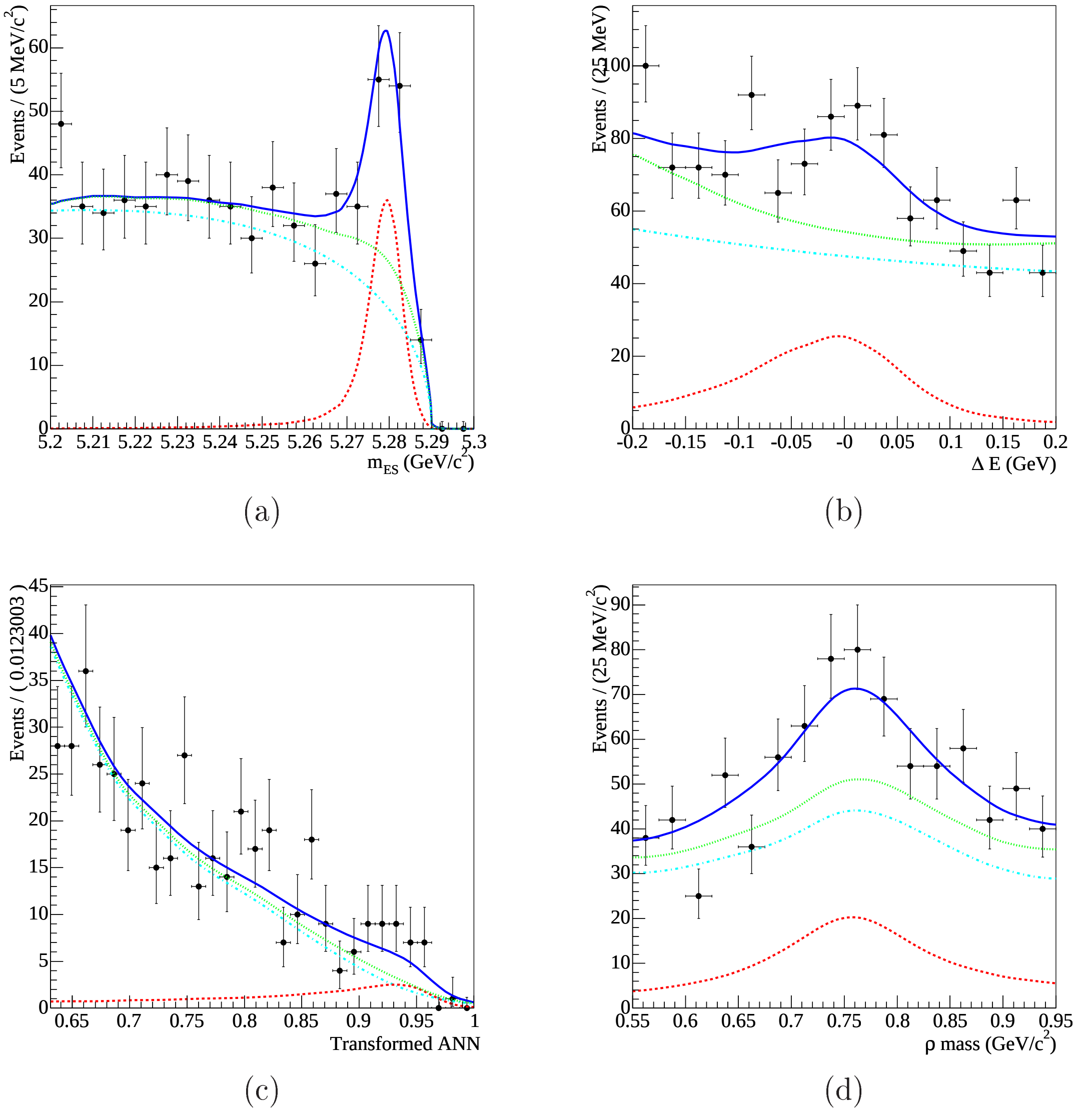


Figure 5.4: Plots of the fit variables with a cut on likelihood ratio for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis. The distributions for m_{ES} (a), ΔE (b), transformed ANN (c), and ρ mass (d) are shown, for candidates passing a cut on the likelihood ratio R (see text). The plots are for the $B^{\pm} \to \rho^{\pm}\pi^{0}$ analysis. In each plot the solid line represents the total PDF, the dotted-dashed line represents the continuum contribution, the dotted line represents the continuum + B background, and the dashed line represents the signal component.

distort the distributions of points in some way. The manner in which this is done is not obvious as the structures of the PDFs are complicated. It is only when the distribution of points over all four variables is considered that one can make a proper judgement.

5.5 Conclusions

$B^{\pm} \to K^{*\pm}\pi^0$ conclusions

Measurements of the charge asymmetry and branching fraction for the decay $B^{\pm} \to K^{*\pm}\pi^0$ ($K^{*\pm} \to K^{\pm}\ \pi^0$) have been made using an extended maximum likelihood fit. The continuum background contamination was floated within the fit and the B–related contamination was estimated using Monte Carlo simulation and its normalisation was fixed in the fit. The results obtained for the branching fraction measurement are multiplied by a factor of 3 to account for the isospin decomposition of the final state, to give a total branching fraction and asymmetry of:

$$
\begin{aligned}
\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^0) &= [6.9\ \pm 2.1\ (Stat.)\ \pm 1.2\ (Syst.)] \times 10^{-6} \\
\mathcal{A}_{CP}(B^{\pm} \to K^{*\pm}\pi^0) &= 0.03\ \pm 0.29\ (Stat.)\ \pm 0.05\ (Syst.).
\end{aligned}
$$

The statistical errors are consistent with errors expected from Monte Carlo studies. The significance of the branching fraction result including systematic errors is 3.8σ showing evidence for this decay. The systematic error of the branching fraction is dominated by the background modelling as well as the uncertainty in the π^0 reconstruction efficiency. The dominant systematic error on the charge asymmetry comes from the uncertainty in the backgrounds' asymmetries.

These experimental results may be compared to the theoretical predictions given in Chapter 1. The prediction of the branching fraction of Neubert and Beneke [30], derived from QCD factorization, is $\mathcal{B}(B^{\pm} \to K^{*\pm}\pi^0) = (3.3\,^{+1.1+1.0+0.6+4.4}_{-1.0-0.9-0.6-1.4}) \times 10^{-6}$ while that of Chiang and Gronau [31], derived from flavour SU(3) symmetry, gives a value between $(15.0^{+3.3}_{-2.8}) \times 10^{-6}$ and $(22.1^{+4.2}_{-5.1}) \times 10^{-6}$. This experimental result favours the former prediction but is consistent within $\sim 2\sigma$ of the lower bound of the latter. It should be noted that the large range of this prediction is due primarily to uncertainties in the angle γ.

The prediction of Beneke and Neubert for the direct *CP* asymmetry in this mode is $(8.7^{+2.1+5.0+2.9+41.7}_{-2.6-4.3-3.4-44.2}) \times 10^{-2}$. Clearly both the experimental result and theoretical prediction are dominated by large errors, and comparison of the two provides no discrepancy.

$B^\pm \to \rho^\pm\pi^0$ conclusions

Measurements of the charge asymmetry and branching fraction for the decay $B^\pm \to \rho^\pm\pi^0$ have been made using an extended maximum likelihood fit. The results obtained

$$\begin{aligned}
\mathcal{B}(B^\pm \to \rho^\pm\pi^0) &= [10.3 \ \pm 1.4\ (Stat.)\ \pm 0.9\ (Syst.)]\ \times 10^{-6} \\
\mathcal{A}_{CP}(B^\pm \to \rho^\pm\pi^0) &= 0.03\ \pm 0.13\ (Stat.)\ \pm 0.02\ (Syst.).
\end{aligned}$$

The previous *BABAR* measurement, using Runs 1 and 2 ($\sim 82\,\mathrm{fb}^{-1}$) was

$$\begin{aligned}
\mathcal{B}(B^\pm \to \rho^\pm\pi^0) &= [10.9 \ \pm 1.9\ (Stat.)\ \pm 1.9\ (Syst.)]\ \times 10^{-6} \\
\mathcal{A}_{CP}(B^\pm \to \rho^\pm\pi^0) &= 0.24\ \pm 0.16\ (Stat.)\ \pm 0.06\ (Syst.).
\end{aligned}$$

The results of the branching fraction measurement are in good agreement. As this analysis and the previous analysis both use the same data, the shift in the central value of the charge asymmetry measurement bears comment.

If the analyses were entirely different such that there was no overlap of events, and the systematic errors were also independent then the errors from the two analysis all add in quadrature to yield $\sqrt{0.13^2 + 0.02^2 + 0.16^2 + 0.06^2} = 0.22$. The difference of 0.21 between the central values is thus within one standard deviation and entirely consistent. This scenario is rather unlikely however and there is a good chance that some events overlap. One could then take the other extreme and conservatively assume the two analyses overlap completely.

To make a proper comparison between the results then one can break down the result for Runs 1-4 into two independent measurements - one for Runs 1-2 and one for Runs 3-4. The estimated measurement for Runs 1-2 is easy - we assume the same central value as the previous analysis and scale the Runs 1-4 error with luminosity. This is calculated to be $\sqrt{211/82} \times 0.13 = 0.21$, such that the Runs 1-2 asymmetry measurement is 0.24 ± 0.21. The asymmetry from Runs 3 and 4 ($129\,\mathrm{fb}^{-1}$) must be (also by luminosity scaling)

$$(0.03 \times 211 - 82 \times 0.24)/129 = -0.10,$$

with a statistical error of $\sqrt{211/129} \times 0.13 = 0.17$. The question is then are the two values -0.10 ± 0.17 and 0.24 ± 0.21 from the two data sets compatible? The significance of the discrepancy is calculated as $(0.24-(-0.10))/\sqrt{0.17^2 + 0.16^2} = 1.5$ standard deviations. However this is the most conservative approach to take in the estimation, and if one accepts that the value of -0.10 depends on the value of the previous measurement (i.e. 0.24) then the errors are correlated and add to 0.38, which gives a significance of less than 1. Thus depending on the degree of overlap one assumes, the degree of compatibility is between 1 and 1.5 standard deviations, and thus entirely consistent.

The change in systematic uncertainty on the branching fraction is primarily due to a refinement of the neutrals correction, which is the dominant source of uncertainty and was larger for the Runs 1 and 2 data. The dominant sources of systematic error on the asymmetry measurement in the previous analysis were from B backgrounds (0.05) and the signal modelling (0.03). Although more is known about the branching fractions of the dominant backgrounds now than for the Runs 1 and 2 analysis, and that the MC simulation is superior, the fundamental differences between the analyses are the most likely factors for accounting for the reduction in the systematic error. In the previous analysis, the ρ mass was used as a variable in their neural network, here it is used directly in the fit. The mass range used previously was 0.4-1.3 GeV/c^2, far wider than the range used here and therefore allowing larger background contributions. Another difference between the analyses was with the tagging information - previously the fit was divided into tagging categories, whereas here tagging information is used in the neural network.

The current world average [17] measurement is $12.0 \pm 2.0 \times 10^{-6}$, making the measurement presented in this thesis superior.

Final conclusions

It is interesting to compare the ratio of the measured branching fractions $B(B^{\pm} \to K^{*\pm}\pi^{0})/B(B^{\pm} \to \rho^{\pm}\pi^{0})$ with the ratio $B(B^{\pm} \to K^{\pm}\pi^{0})/B(B^{\pm} \to \pi^{\pm}\pi^{0})$:

$$\frac{B(B^{\pm} \to K^{*\pm}\pi^{0})}{B(B^{\pm} \to \rho^{\pm}\pi^{0})} = \frac{6.9 \pm 2.4}{10.3 \pm 1.7} = 0.7 \pm 0.3 \tag{5.3}$$

$$\frac{B(B^{\pm} \to K^{\pm}\pi^{0})}{B(B^{\pm} \to \pi^{\pm}\pi^{0})} = \frac{12.1 \pm 0.8}{5.5 \pm 0.6} = 2.2 \pm 0.3 \tag{5.4}$$

where any systematic cancellations associated with the measurements have been ignored. The penguin contribution to $K\pi$-like modes from CKM arguments are already expected to be significant - the penguins are of order $[V_{tb}^{*}V_{ts}] \sim \lambda^{2}$ compared to the tree contribution which is of order $[V_{ub}^{*}V_{us}] \sim \lambda^{4}$ - while the $B^{\pm} \to \pi^{\pm}\pi^{0}$ is a $\Delta I = 3/2$, tree-dominated process. The fact that (5.3) is much smaller than (5.4) is suggestive that the penguin contributions to the $K^{*\pm}\pi^{0}$ VP process are smaller than for the PP, where P and V denote pseudoscalar and vector states respectively. From QCD factorization, this is entirely what is expected due to the behaviour of the Wilson coefficients prescribing the relative strengths of the contributing amplitudes in the effective weak Hamiltonian [30].

In [30], the default theoretical prediction for the ratio Eq.(5.3) is 0.24. However, interestingly there are four additional predictions made under differing sets of theoretical parameter scenarios. One of which, named 'S4', makes a prediction of the ratio Eq.(5.3) as being 0.63. This scenario assumes certain particular values for the strange-quark mass, the ratio $|V_{ub}/V_{cb}|$, and the coefficients of the decay operators in the effective Hamiltonian (amongst others). The value of the ratio Eqn.(5.3) presented here gives closest agreement with this 'S4' QCD factorization parameter scenario.

References

[1] T. D. Lee and C. N. Yang, "Parity Nonconservation and a Two-Component Theory of the Neutrino", *Phys. Rev.* **105** (1957) 1671.

[2] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the $2\,\pi$ Decay of the K_2^0 Meson", *Phys. Rev. Lett.* **13** (1964) 138.

[3] `http://www.slac.stanford.edu/BFROOT/`.

[4] `http://belle.kek.jp`.

[5] *BABAR* Collaboration, B. Aubert *et al.*, "Observation of *CP* violation in the B^0 meson system", *Phys. Rev. Lett.* **87** (2001) 091801, `hep-ex/0107013`.

[6] BELLE Collaboration, K. Abe *et al.*, "Observation of Large *CP* violation in the neutral *B* meson system", *Phys. Rev. Lett.* **87** (2001) 091802, `hep-ex/0107061`.

[7] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interactions", *Prog. Theor. Phys.* **49** (1973) 652.

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

[9] A. Salam and J. Ward *Phys. Lett.* **13** (1964) 168.

[10] S. Glashow *Nucl. Phys* **22** (1961) 579.

[11] A. D. Sakharov *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32.

[12] S. Weinberg, "The Quantum Theory of Fields", vol. 1, Cambridge University Press, 1995.

[13] I. I. Bigi and A. I. Sanda, "CP Violation", pp. 38–50. Cambridge University Press, 2000.

[14] G. Arfken and H. Weber, "Mathematical methods for physicists", Academic Press, 2000.

[15] S. Eidelman *et al.*, "Review of Particle Physics", *Phys. Lett. B* **592** (2004).

[16] L. Wolfenstein, "Parameterization of the Kobayashi-Maskawa matrix", *Phys. Rev. Lett.* **51** (1983) 1945.

[17] Heavy Flavour Averaging Group (HFAG), "Averages of b-hadron Properties as of Summer 2004", `hep-ex/0412073`. `http://www.slac.stanford.edu/xorg/hfag/rare/index.html`.

[18] *BABAR* Collaboration, B. Aubert *et al.*, "Direct CP violating asymmetry in $B^0 \to K^+\pi^-$ decays", *Phys. Rev. Lett.* **93** (2004) 131801, `hep-ex/0407057`.

[19] BELLE Collaboration, Y. Chao *et al.*, "Evidence for direct CP violation in $B^0 \to K^+\pi^-$ decays", *Phys. Rev. Lett.* **93** (2004) 191802, `hep-ex/0408100`.

[20] P. Harrison and H. Quinn, eds, "The *BABAR* physics book", SLAC-R-504, 1998.

[21] M. Gronau, "Penguins in B decays", *AIP Conf. Proc.* **618** (2002) 256–265.

[22] M. Gronau, "Large penguin effects in the *CP* Asymmetry of $B^0 \to \pi^+\pi^-$", *Phys. Lett.* **B 300** (1993) 163.

[23] M. Gronau and D. London, "Isospin analysis of CP asymmetries in B decays", *Phys. Rev. Lett.* **65** (1990) 3381.

[24] A. E. Snyder and H. Quinn, "Measuring *CP* asymmetry in $B \to \rho\pi$ decays without ambiguities.", *Phys. Rev.* **D 48** (1993) 2139.

[25] M. Gronau, J. Rosner, and D. London, "Weak coupling phase from decays of charged B mesons to πK and $\pi\pi$.", *Phys. Rev. Lett.* **73** (1994) 21.

[26] M. Gronau, J. Rosner, and D. London, "Decays of B mesons to two light pseudoscalars.", *Phys. Rev.* **D 50** (1994) 4529.

[27] T. Yoshikawa, "Possibility of large electroweak penguin contribution in $B \to K\pi$ modes.", *Phys. Rev.* **D 68** (2003) 1.

[28] A. Hoecker *et al.*, "`http://www.slac.stanford.edu/xorg/ckmfitter/ckm_wellcome.html`.".

[29] D. Atwood and A. Soni, "Possibility of large direct *CP* violation in $B \to K\pi$ -like modes because of long distance rescattering effects and implications for the angle γ.", *Phys. Rev.* **D 58** (1998) 1.

[30] M. Beneke and M. Neubert, "QCD factorization for $B \to PP$ and $B \to PV$ decays", *Nucl. Phys.* **B 675** (2003) 333.

[31] C. W. Chiang, M. Gronau, Z. Luo, J. Rosner, and D. Suprun, "Charmless $B \to VP$ decays using flavor SU(3) symmetry", *Phys. Rev.* **D 69**, (2004) 034001.

[32] CLEO Collaboration, P. Jessop, C *et al. Phys. Rev. Lett.* **85** *2881* (2000).

[33] R. H. Dalitz, "On the Analysis of τ Meson Data and the Nature of the τ Meson", *Phil. Mag.* **44** (1953) 1068.

[34] "`http://www.slac.stanford.edu/grp/ad/pep-ii/stats/PEPII_lum_record.pdf`.".

[35] *BABAR* Collaboration, B. Aubert *et al.*, "The *BABAR* Detector", *Nucl. Instrum. Meth.* **A479** (2002) 1, `hep-ex/0105044`.

[36] P. Dauncey, Private communication.

[37] D. Lange and A. Ryd, "The EvtGen Particle Simulation Package", *Nucl. Instrum. Meth.* **A462** (2001) 152. `http://hep.ucsb.edu/people/lange/EvtGen/`.

[38] T. Sjostrand, "PYTHIA 5.7 and JETSET 7.4 Physics and Manual", *CERN-TH-7112-93-REV* (1995), `hep-ph/9508391`.

[39] S. Agostinelli *et al.*, "GEANT4 – a simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250.

[40] J. D. Bjorken and S. J. Brodsky *Phys. Rev.* **D 1** (1970) 1416.

[41] G. Fox and S. Wolfram *Nucl. Phys.* **B 149** (1979) 413.

[42] V. Tisserand and D. Payne, "`http://www.slac.stanford.edu/BFROOT/www/Physics/Analysis/AWG/Neutrals%/validation/Summer2004/summer2004.html`.".

[43] M. Allen, M. Naisbit, and A. Roodman, "A Study of π^0 Efficiency", *BABAR Analysis Document 870, Version 3* (2004).

[44] S. Spanier and G. Mancinelli, "Kaon Selection at *BABAR*", *BABAR Analysis Document 116, Version 1* (2002).

[45] C. Roat, "`http://www.slac.stanford.edu/BFROOT/www/Physics/Tools/Pid/Protons/description.html`.".

[46] R. Fisher, "The Use of Multiple Measures in Taxonomic Problems.", *Ann. Eugenics* **7** (1936) 179.

[47] P. C. Mahalanobis, *Proc. Natl. Institute of Science of India* **2** (1936) 49.

[48] T. J. Harrison, "Searches for B^0 Decays to $\pi^+\pi^-\pi^0$", *BABAR Analysis Document 159, Version 11.*

[49] W. Y. Huang and R. P. Lippmann, "Neural net and traditional classifiers", *Neural Information Processing Systems* (1987) 387. `http://books.nips.cc/nips00.html`.

[50] T. Rognvaldsson, "Langevin Updating in Multilayer Perceptrons", *Neural Comp 6* **5** (1994).

[51] J. Beringer, “BTagger - A Multivariate Tagging Algorithm with Categories Based on the Physics of the Btag Decay.” *BABAR Analysis Document 317, Version 4* 2002.

[52] C. Peterson and T. Rognvaldsson, “JETNET 3.0 - A Versatile Artificial Neural Network Package”, *CERN_TH* **7135/94** (1993).

[53] K. Phillip *et al.* `http://www.physics.ohio-state.edu/~catutza/root_to_jetnet/rtj.html`.

[54] S. Menke, “Calibration of the BaBar Electromagnetic Calorimeter with π^0 s.” *BABAR* Note 528, Version 1, 2001

[55] R. J. Barlow, “Statistics. A guide to the use of statistical methods in the physical Sciences”, Wiley, 1997.

[56] F. James, “CERN Program Library Long Writeup D506, MINUIT Reference Manual, Version 94.1”,. Computing and Networks Division, CERN Geneva, Switzerland, 1994.

[57] W. Verkerke and D. Kirkby, “The RooFit Toolkit for Data Modelling.” `http://roofit.sourceforge.net`.

[58] R. Brun *et al.*, “ROOT”, `http://root.cern.ch`.

[59] K. Cranmer, “Kernel Estimation in High-Energy Physics”, *Comput. Phys. Commun.* **136** (2000) 198, hep-ex/0011057.

[60] *ARGUS* Collaboration, H. Albrecht *et al. Phys. Lett. B* **241** (1990) 278.

[61] *BABAR* Collaboration, B. Aubert *et al.*, “Measurement of the branching fractions and charge asymmetries in $B^{\pm} \to \rho^{\pm}\pi^0$ and $B^{\pm} \to \rho^0\pi^{\pm}$ decays, and search for $B^0(\overline{B}^0) \to \rho^0\pi^0$”, *Phys. Rev. Lett.* **93** (2004) 051802.

[62] R. Aleksan *et al.*, “Measurement of Branching Fractions and CP-Violating Asymmetries in $B^0 \to \rho^+ h^-$.” 2003.

[63] A. Mihalyi *et al.*, "Study of Branching Fraction and CP-Violating Asymmetry in $B^{\pm} \to \rho^{\pm}\pi^{0}$." 2003.

[64] G. Graziani, "Direct CP violation searches and related branching fractions", FPCP. 2004.

[65] S. Menke, "Treatment of correlated systematic errors", *BABAR Analysis Document 387, Version 1* (2002).

[66] B. Aubert *et al.*, "Branching Fractions and CP Asymmetries in $B^0 \to \pi^0\pi^0$, $B^{\pm} \to \pi^{\pm}\pi^0$ and $B^{\pm} \to K^{\pm}\pi^0$ Decays and Isospin Analysis of the $B \to \pi\pi$ System", Submitted to *Phys. Rev. Lett* (2004) `hep-ex/0412037`.

[67] J. Malcles *BABAR Analysis Document 952* (2004).

[68] D. Payne and V. Tisserand, `http://www.slac.stanford.edu/BFROOT/www/Physics/Analysis/AWG/Neutrals/`.

[69] T. Allmendinger, `http://www.slac.stanford.edu/BFROOT/www/Physics/TrackEfficTaskForce/TrackingTaskForce-2004.html`.

[70] G. Mancinelli, `http://www.slac.stanford.edu/BFROOT/www/Organization/CollabMtgs/2001/detDec2001/Wed1a/giampi.pdf`.

[71] T. Brandt and D. Aston `http://www.slac.stanford.edu/BFROOT/www/Physics/Tools/Pid/pid.html`.

[72] C. Hearty, "Measurement of the Number of $\Upsilon(4S)$ Mesons Produced in Run 1 (B Counting)" *BABAR Analysis Document 314* (2001).