# Measurement of B -> D Form Factors in the Semileptonic Decay B^0 -> D* 1 nu at BABAR 

M. S. Gill<br>Stanford Linear Accelerator Center<br>Stanford University<br>Stanford, CA 94309

SLAC-Report-794

Prepared for the Department of Energy
under contract number DE-AC02-76SF00515

Printed in the United States of America. Available from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

This document, and the material and data contained therein, was developed under sponsorship of the United States Government. Neither the United States nor the Department of Energy, nor the Leland Stanford Junior University, nor their employees, nor their respective contractors, subcontractors, or their employees, makes an warranty, express or implied, or assumes any liability of responsibility for accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use will not infringe privately owned rights. Mention of any product, its manufacturer, or suppliers shall not, nor is it intended to, imply approval, disapproval, or fitness of any particular use. A royalty-free, nonexclusive right to use and disseminate same of whatsoever, is expressly reserved to the United States and the University.

Measurement of $B \rightarrow D$ Form Factors in the Semileptonic Decay $B^{0} \rightarrow D^{*} \ell \nu$ at BaBar
by
Mandeep Singh Gill
B.A. (University of California Berkeley)

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy
in

Physics
in the

GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:
Professor Robert Jacobsen, Chair
Professor Marjorie Shapiro
Professor Eugene Haller

The dissertation of Mandeep Singh Gill is approved:

| Chair | Date |
| :---: | :---: |
|  |  |
|  | Date |
|  |  |

University of California at Berkeley

Fall 2004

Abstract<br>Measurement of $B \rightarrow D$ Form Factors in the Semileptonic Decay $B^{0} \rightarrow D^{*} \ell \nu$ at BaBar<br>by<br>\section*{Mandeep Singh Gill}<br>Doctor of Philosophy in Physics<br>University of California at Berkeley<br>Professor Robert Jacobsen, Chair

We present here the results of a measurement of the three semileptonic form factors involved in the decay $B^{0} \rightarrow D^{*} \ell \nu$, where $\ell$ is one of the two light charged leptons (i.e. an electron or muon - though the final results in this work are determined only for $\ell$ $=$ electron). This measurement uses the Babar 2000-2002 data set, which is altogether approximately $85 \times 10^{6} B \bar{B}$-pairs in $78 \mathrm{fb}^{-1}$ of integrated luminosity. The $D^{*+}$ was reconstructed in the channel $D^{*+} \rightarrow D^{0} \pi^{+}$, and the $D^{0}$ in the channel $D^{0} \rightarrow K^{-} \pi^{+}$. This analysis was based ultimately on $\sim 16,386$ reconstructed events with an estimated background contamination of $\sim 15 \%$. The method of the measurement was to perform a unbinned maximum likelihood fit in the four kinematic variables that describe the decay for the three form factor parameters $R_{1}, R_{2}$, and $\rho^{2}$. The results obtained for the form
factor ratios are $R_{1}=1.328 \pm 0.055 \pm 0.025 \pm 0.025$ and $R_{2}=0.920 \pm 0.044 \pm 0.020 \pm 0.013$ for the ratios and $\rho^{2}=0.769 \pm 0.039 \pm 0.019 \pm 0.032$ for the form factor slope. The errors given are statistical, Monte Carlo statistical and systematic respectively.

I dedicate this dissertation to Santosh Auntie, Minna, Ginny, Karan, Mathai, and H.Thaiji. All of whom escaped this mortal coil during the pursuit of this work.

Gone from this Earth, but not ever forgotten.
WJKK, SWJKF.

## Contents

List of Figures ..... vii
List of Tables ..... xv
1 Introduction and Motivations ..... 1
1.1 Standard Model ..... 2
1.1.1 Elementary Particles - quarks, leptons and forces ..... 3
1.1.2 Composite QCD Particles - Baryons and Mesons ..... 8
1.1.3 Particle decays ..... 11
1.1.4 CP violation ..... 12
1.1.5 Electroweak Interactions ..... 15
1.2 The Role of Form Factors in Semileptonic Decays ..... 17
1.2.1 Heavy Quark Effective Theory (HQET) ..... 19
1.3 Determination of CKM element $V_{c b}$ ..... 20
1.4 Determination of CKM element $V_{u b}$ ..... 21
1.5 This Measurement ..... 22
2 Analysis Method Overview ..... 23
3 Formalism ..... 28
3.1 Quark-level Description ..... 28
3.2 Angular Kinematic Variables ..... 30
3.3 Four Dimensional Probability Distribution Function ..... 32
3.3.1 HQET helicity amplitudes ..... 34
3.3.2 Simplified PDF ..... 34
3.4 Helicity Amplitudes and PDF Factors ..... 35
3.5 Contributions of $H_{[+, 0,-]}^{2}$ ..... 38
3.6 Form Factor Predictions ..... 41
4 BABAR Detector Overview ..... 44
4.1 Detector Overview ..... 44
4.2 The Positron-Electron Project (PEP) II ..... 46
4.2.1 Asymmetric Collider ..... 47
4.2.2 Accelerating Particles ..... 47
4.3 The Silicon Vertex Tracker (SVT) ..... 49
4.4 The Drift Chamber (DCH) ..... 52
4.5 The Detector of Internally Reflected Cerenkov Light (DIRC) ..... 55
4.6 The Electromagnetic Calorimeter (EMC) ..... 58
4.7 The Instrumented Flux Return (IFR) ..... 61
4.8 The Trigger System ..... 62
4.8.1 Level 1 Trigger ..... 63
4.8.2 Level 3 Trigger ..... 65
5 Event Selection Description for $B^{0} \rightarrow D^{*} \ell \nu \quad$ Candidates ..... 68
5.1 BABAR Monte Carlo (MC) Description ..... 68
5.1.1 Generator Level MC - EvtGen ..... 69
5.1.2 Full MC - Simulation Production 4 ..... 69
$5.2 \quad B^{0} \rightarrow D^{*} \ell \nu$ Event Selection ..... 71
5.2.1 Initial Data/MC Processing ..... 71
5.2.2 Decay Channels Considered ..... 71
5.2.3 Data and MC Set ..... 73
5.2.4 Backgrounds ..... 74
5.2.5 Reconstruction track lists ..... 76
5.2.6 Composition Tools ..... 79
5.2.7 Particle Identification ..... 81
5.2.8 Tagbit selection ..... 90
5.2.9 Analysis Cuts ..... 92
5.2.10 Cut efficiency ..... 114
5.2.11 Summary of Final Cuts ..... 121
6 Fitting Method ..... 125
6.1 Motivation for Moments Expansion Fitting Method ..... 125
6.1.1 Break PDF into Products of Helicity Amplitudes and Angular Pieces ..... 127
6.1.2 $\zeta$ parameter product basis ..... 128
6.2 Moments With a Realistic Detector: Acceptance/Efficiency Effects ..... 138
6.2.1 $\xi$ parameters ..... 138
6.2.2 Expression for $\bar{\epsilon}$ ..... 141
6.2.3 $\xi$ Correction Factors from Sums Over Events ..... 142
6.2.4 Fitting to Detector-corrected Sample ..... 144
6.2.5 Self-calibration Theorem ..... 145
6.3 SP4 Moments Method Validation ..... 147
6.3.1 SP4 calibration on SP4 fit ..... 147
6.3.2 Diamond Frame ..... 149
6.4 Conceptual Description of Background Subtraction Method ..... 153
6.5 Full MC Validation of DUBS Method ..... 156
6.5.1 SP4 calibration on SP4 fit ..... 156
6.6 Bias Map Method for Correcting for Resolution ..... 157
6.6.1 Bias Map Correction Error ..... 158
6.6.2 Numerical Values of Bias Map Correction ..... 159
6.7 Fit results ..... 159
6.8 Goodness of Fit Evaluation Method ..... 161
6.8.1 GOF on Reweighted Samples ..... 161
6.8.2 GOF Results on our Dataset ..... 161
7 Relevant MC and Data Plots ..... 168
7.1 Signal vs. Backgrounds in several basic variables ..... 168
7.2 MC vs. Data Plots and Yields ..... 170
7.2.1 EvtGen Plots ..... 173
7.3 Relevant Full MC (SP4) vs. Data and EvtGen Plots ..... 174
7.3.1 Primary Signal Region SP4 vs. Data Plots ..... 174
7.3.2 EvtGen kinematic variable Plots with varying parameters ..... 177
7.3.3 Lepton momentum in SP4 vs. Data and EvtGen ..... 180
8 Systematic Error Estimation ..... 181
8.1 Kinematic Smearing ..... 182
8.1.1 Bias Map Correction and Errors with Multiple Parameters ..... 182
8.1.2 How resolution enters the calculation ..... 185
8.1.3 Results From SP4 ..... 186
8.2 Intrinsic Errors to Efficiency Moment Method ..... 187
8.2.1 From variation in parameters for calibration file ..... 187
8.2.2 Efficiency Moment Statistical Error (EMSE) With Correlated Param- eters ..... 189
8.2.3 EMSE error - Full MC Results ..... 194
8.3 Intrinsic DUBS Method Errors ..... 196
8.3.1 DUBS error - Analytic Results on Generator-level MC Samples ..... 202
8.3.2 DUBS error - Full MC Results ..... 202
8.4 Detector Related Errors ..... 204
8.4.1 Tracking ..... 204
8.4.2 Slow Pion Tracking ..... 207
8.4.3 PID ..... 209
8.4.4 $\delta m$ Width difference ..... 211
8.4.5 Error due to FSR ..... 212
8.4.6 Peaking Backgrounds ..... 215
8.4.7 Other semileptonic channel contributions ..... 215
8.4.8 Combinatoric $\delta m$ background difference in shape ..... 219
8.5 Final Errors ..... 221
9 Conclusions ..... 224
Bibliography ..... 226
A Notation Used in Text ..... 231
A. 1 Colors ..... 231
A. 2 Standard Form Factor Parameter Notation ..... 232
B Including CP violation and massive lepton terms in Fitting PDF ..... 233
C EvtGen and RooFit PDF Validation ..... 236
C. 1 EvtGen and RooFit Validation from Kinematic Variable Distribution Com- parison ..... 236
C. 2 Showing Equivalence of PDF by Fitting ..... 238
C. 3 Ntuple production ..... 242
D EvtGen Moments Method Validation ..... 246
D. 1 Fix calibration on one file and fit to files with varying parameters ..... 246
D. 2 First test of acceptance correction method with cuts ..... 247
D. 3 Tests varying fit samples ..... 249
E SP4 Fits by separate $D^{0}$ and lepton modes ..... 251
F YFrame ..... 257
G Avoiding Spurious PDF Zeros for Background by PDF $\chi$-Averaging ..... 261
H EvtGen Validation of DUBS Method ..... 263
I Examples of need for bias map correction ..... 265
I. 1 Simple one dimensional example of need for bias map correction ..... 265
I. 2 Extra example of need for bias map correction ..... 270
J Extra Goodness-of-Fit Validations ..... 275
J. 1 GOF on EvtGen samples ..... 275
J. 2 GOF on SP4 samples ..... 277
J. 3 GOF on Data vs. MC background control samples ..... 278
K Extra MC vs. Data Plots ..... 280
K. 1 Signal and Sideband Region SP4 vs. Data Plots By Lepton Type ..... 281
K. 2 Signal and Sideband Region SP4 vs. Data Plots By $D^{0}$ Mode ..... 283
K. 3 Angular Kinematic Variables with cuts on other kinematic variables ..... 287
K. 4 SP4 vs. Data 2D Kinematic Variable Correlation Plots ..... 289
K. 5 Slow pion momentum in SP4 vs. Data and EvtGen ..... 294
K. 6 Data/MC Control Plots ..... 296
L EvtGen EMSE and DUBS errors validation ..... 299
L. 1 EMSE error - Analytic EvtGen Results ..... 299
L.1.1 EMSE error - Comparison of EvtGen Results to ToyMC ..... 301
L.1.2 EMSE error - Comparison of Analytic Calibration Shift Error to Calculated Error ..... 301
L. 2 DUBS error - Analytic Results on Evtgen Samples ..... 305
L.2.1 DUBS error - Comparison of Analytic Results to Evtgen ToyMC ..... 307
M $M_{D^{*}}-M_{D^{0}}(\delta M)$ Systematic Error Details ..... 309
N FSR SystErr Details ..... 311

## List of Figures

3.1 Quark level diagram of $B \rightarrow D^{*} \ell \nu$ decay. ..... 28
3.2 Meson level diagram of $B \rightarrow D^{*} \ell \nu$ decay (for the decay we are analyzing, the $B$ is neutral, the $D^{*}$ charged, as is the $\pi$, and the $D$ daughter meson is neutral). ..... 31
3.3 Various helicity component contributions to the PDF (plots made with the CLEO values of the form factors) ..... 39
3.4 Minimal ( $w=1.000$ ) configuration ..... 39
3.5 Maximal $(w=1.504)$ configuration. ..... 40
4.1 BaBar detector; the standard coordinate system is defined with $+z$ point- ing along the electron beam direction (to the right in this figure), and $\theta$ and $\phi$ are then the standard polar and azimuthal angles with respect to this axis. ..... 45
4.2 SLAC linac and storage rings. ..... 46
4.3 SVT profile; note five-layer concentric shell structure. ..... 49
4.4 SVT resolution for data and MC as a function of incident angle. ..... 51
4.5 Side view of the Drift Chamber layout, the high energy electrons are trav- eling to the right. As with the SVT, the interaction point (IP) is located closer to the high energy side to improve solid angle coverage as the par- ticles spray preferentially out towards the right hand (low energy) side. Note dimensions are given in millimeters in this diagram. ..... 53
4.6 Drift Chamber resolution vs track distance from wire at 1961V. ..... 55
$4.7 \mathrm{dE} / \mathrm{dx}$ vs p from Drift Chamber; the units on the vertical axis are propor- tional to energy (loss) per unit length. ..... 56
4.8 DIRC structure, emphasizing quartz bars, support framework, and Stand- off Box. ..... 57
4.9 DIRC operating principle, showing total internal reflection of the photons, and their detection in PMT's in the Standoff Box. ..... 58
4.10 DIRC single photon resolution and photons per track. ..... 59
4.11 EMC position and energy resolution (the $\sigma^{\prime}$ s are the calibration-derived parameters that determine the resolution of the detector). ..... 60
4.12 Left: Elements which make up the Level 1 trigger. The TSF (Track Segment Finder) sends track segment information to the BLT (Binary Link Tracker) and PTD ( $P_{t}$ Discriminator). The BLT, PTD, EMT and IFT all send their respective bit map information to the GLT which decides if a Level 1 accept should be issued. Right: DCT efficiency plot. The A and B tracks turn on curves are determined by the minimum $p_{t}$ required to reach half way and all the way through the Drift Chamber. The A' track turn on is programmable, currently set at $800 \mathrm{MeV} / c$.
4.13 L3 event display. The blue circles are track segments from the TSF boards. The red lines label reconstructed charged tracks and the red towers show energy deposits reported by the EMT.
5.1 Distributions of (a) total momentum, (b) transverse momentum, (c) cosine of polar angle, and (d) azimuthal angle (all measured in the lab frame) for the following lists: ChargedTracks, GoodTracksVeryLoose, GoodTracksLoose, GoodTracksTight. The plots are made from a typical run (number 12917), normalized to display tracks/event/bin. The tagbit (see Section 5.2.8) used for generating these plots is called isPhysics, this is a very loose tagbit which essentially only requires an event to have more than two charged tracks to remove $e^{+} e^{-}$and $\mu^{+} \mu^{-}$events [26].
5.2 Left, the distribution of Cerenkov angle versus momentum for different particle types (the top band is from pions, the next one is from kaons). Right, the kaon/pion separation in units of $\sigma$.89
5.3 A plot of $D^{0}$ masses for all $D^{0}$ candidates (red dashed) and MC matched candidates (solid blue) for the four $D^{0}$ decay modes. The final cut values are shown by the black vertical lines.
5.4 Top: Plots of $D^{0}$ vertex probabilities for all $D^{0}$ candidates (left) and MC matched candidates (right). Bottom left: This plot shows the effect of the D0Vtx cut on all events - the red dashed curve is all events while the solid blue curve is events passing the $D^{0}$ vertex cut. Bottom right: This shows the effect of the cut on MC-matched events. The red dashed curve is all MC-matched events before the cut, the solid blue curve is MC-matched events after the cut.
5.5 The effect of applying Kaon ID to the the $D^{0}$ samples. The red dashed curve shows all events, the solid blue curve is all events which pass Kaon ID. The $K \pi$ mode uses the "not a pion" requirement which has virtually no effect at all, the $K^{-} \pi^{+} \pi^{+} \pi^{-}$(left) and $K \pi^{+} \pi^{0}$ (right) modes use the "tight kaon" particle ID, these two modes are shown above.
5.6 The effect of applying Kaon ID to the the $D^{0}$ samples on MC-matched events. The red dashed curve shows all MC-matched events, the solid blue curve shows the MC-matched events remaining after the Kaon ID cut ( $K^{-} \pi^{+} \pi^{+} \pi^{-}$mode (left) and $K \pi^{+} \pi^{0}$ mode (right)).
5.7 Top: $D^{0} \rightarrow K \pi \pi^{0}$ Dalitz distribution at the generator level without convolving the distribution with resolution functions. Bottom: The region which falls inside the typical cut based region of $\left|m\left(\pi^{+} \pi^{-}\right)-m(\rho)\right|<250$ $\mathrm{MeV} / c^{2}$ and $\left|\cos \theta_{K \pi^{+}}^{*}\right|>0.4$ in the $\pi^{+} \pi^{-}$rest frame. . . . . . . . . . . . .
$\begin{array}{ll}5.8 & \text { Left: } K_{S} \text { mass plot (for the currently unused } D^{0} \rightarrow K_{S} \pi \pi \text { mode), the } \\ \text { dashed red curve is all events, solid blue curve is MC-matched events. }\end{array}$
5.8 Left: $K_{S}$ mass plot (for the currently unused $D^{0} \rightarrow K_{S} \pi \pi$ mode), the
dashed red curve is all events, solid blue curve is MC-matched events. Right: $\pi^{0}$ mass plot (for the $D^{0} \rightarrow K \pi \pi^{0}$ mode), the dashed red curve is all events, the solid blue curve is MC-matched events. Top: The left plot shows all $\pi^{0}$ events, the right plot shows all MC-matched $\pi^{0}$ events. The vertex cut applied on $\pi^{0}$ is $\operatorname{prob}\left(\chi^{2}\right)>0.01$ (the first bin on both plots). Bottom left: shows all events (red dashed) and events passing the vertex cut (solid blue). Bottom right: the red dashed curve is all MCmatched $\pi^{0}$ events, the solid blue curve is all MC-matched events which pass the vertex cut.
5.10 Left: the momentum distribution of the slow pion. The red dashed curve represents all $D^{*}$ candidate, the solid blue curve is MC-matched $D^{* \prime}$ s. Right: Same as figure on the left except the slow pion candidate is required to have at least 5 DCH hits.
5.11 The momentum of the $D^{*}$ in the center of mass reference frame. The red dotted curve represents all $D^{*}$ candidates, the blue dashed curve is all MCmatched $D^{*}$ candidates while the solid green curve shows all MC-matched $D^{*} \ell \nu$ candidates (this shows especially the effectiveness of the upper limit cut).105
5.12 The vertex quality cut applied to the $D^{*} l$ vertex. Top: The left plot represents all events while the right plot corresponds to MC matched events. Bottom: These are $\delta m$ plots. The left shows all events (red dashed) and events passing the cut (solid blue) while the right shows MC-matched events (red dashed) and MC-matched events passing the vertex cut (solid blue).
5.13 The momentum of the lepton candidate in the center of mass frame. The red dashed curve shows all candidates while the solid blue curve shows MC-matched $D^{*} \ell \nu$ candidates.
5.14 Left: the red dashed curve shows all events while the solid blue curve shows all events passing the good lepton cut (i.e., passes very tight lepton ID). Right: These are $\delta m$ plots. The red dashed curve shows all MCmatched events and the solid blue curve shows all MC-matched events passing the good lepton cut.

$$
\begin{aligned}
& 5.15 \text { The distribution of the thrust variable broken down by different types } \\
& \text { of backgrounds. Upper left: the black dashed line (topmost line) shows } \\
& \text { the thrust distribution after all } D^{0} \text { and } D^{*} \text { cuts have been made (includ- } \\
& \text { ing the } D^{*}-D^{0} \text { mass cut). The solid blue line shows MC-matched } D^{*} \ell \nu \\
& \text { events, the red dashed (lowest line) shows the combinatoric background (it } \\
& \text { is modest because of the mass cut). The other plots show the ccbar, flipped } \\
& \text { (direction-reversed, momentum magnitude kept constant)/cascade (sec- } \\
& \text { onadary) lepton and charged } B \text { plots. . . . . . . . . . . . . . . . . . . } 111
\end{aligned}
$$

5.16 A plot of $\cos \theta_{B, D^{*} l}$ for all events, MC-matched $D^{*} \ell \nu$ signal events and flipped lepton events (i.e. a real $D^{*}$ from one $B$ and a real lepton from the other), after all cuts except for this one $\left(\cos \theta_{B, D^{*} l}\right)$. ..... 112
5.17 A plot of $\cos \theta_{D^{*} l}$, not cut on for this analysis. ..... 113
5.18 The red dashed curve shows all $D^{*} l$ candidate events with no cuts applied. The blue dotted curve shows all real $D^{* \prime}$ in this sample while the solid black curve shows all real $D^{*} \ell \nu$ events in this sample (shown for the four $D^{0}$ decay modes). ..... 117
5.19 The red dashed curve shows all $D^{*} l$ candidate events with only $D^{0}$ cuts applied. The blue dotted curve shows all real $D^{* \prime}$ s in this sample while the solid black curve shows all real $D^{*} \ell \nu$ events in this sample (shown for the four $D^{0}$ decay modes). ..... 118
5.20 The red dashed curve shows all $D^{*} \ell \nu$ candidate events with all $D^{*} \ell \nu$ cuts applied. The solid blue curve shows all real $D^{*} \ell \nu$ events in this sample (shown for the four $D^{0}$ decay modes). ..... 119
5.21 Red dashed shows all MC-matched events prior to cuts, solid blue shows MC-matched events after all cuts applied. This plot shows the efficiency for finding $D^{*} \ell \nu$ events in each decay mode. ..... 120
6.1 Schematic representation of the origin of the term "Diamond Frame" ..... 151
6.2 Side view of $\cos \theta_{B Y}$ reconstruction. The points at $\phi_{B Y}=0$ and $\pi$ are in the $D^{*}-\ell$ plane. The points at $\pm \pi / 2$ are out of the plane. ..... 152
6.3 Data/MC comparison for $w$-distribution (for this and all following plots, red points are data, and green are MC). Top uses CLEO parameters and bottom is re-weighted to our parameters. ..... 163
6.4 Data/MC comparison for $\cos \theta_{l}$-distribution. Top use CLEO parameters and bottom is re-weighted to our parameters. ..... 164
6.5 Data/MC comparison for $\cos \theta_{V}$-distribution. Top uses CLEO parameters and bottom is re-weighted to our parameters. ..... 165
6.6 Data/MC comparison for $\chi$-distribution. Top uses CLEO parameters and bottom is re-weighted to our parameters. ..... 166
6.7 Data/MC comparison for $p_{l}^{*}$-distribution. Top uses CLEO parameters and bottom is re-weighted to our parameters. ..... 167
7.1 This plot shows the contribution of the signal and the different background types to the deltaM distribution.169
7.2 This plot shows the contribution of the signal and the different background types to the deltaM distribution in the signal window. ..... 169
7.3 This plot shows the contribution of the signal and the different background types to the $\cos \theta_{B Y}$ distribution. ..... 170
7.4 Kinematic Variables in the signal window for both leptons. Upper left is $w$, upper right $\cos \theta_{\ell}$, lower left $\cos \theta_{V}$, and lower right is $\chi$. Data is solid red and MC is dashed blue. This is the pattern that all following plots will follow, though they will be smaller. ..... 175
7.5 Kinematic Variables in the sideband region for both leptons. ..... 176
7.6 Kinematic Variable ratios in the sideband region for both leptons, the $\frac{\chi^{2}}{\text { d.o.f. }}$ of $0.92,1.32,1.03$, and 1.57 (for $w, \cos \theta_{\ell}, \cos \theta_{V}, \chi$ ) indicate the amount of agreement in the combinatorial background between Data and MC. . . . . ..... 176
7.7 Kinematic Variables for EvtGen produced files, the legend uses the short- hand notation described in the text. These are the XLL plots. ..... 177
7.8 These are the LXL plots. ..... 178
7.9 These are the LLX plots. ..... 179
7.10 Left: Lepton momentum in the signal window, Right: ratio of data to MC of these histograms. The clear upward trend in the ratio indicates that the form factors in the data are not exactly those in the MC, since the form factors affect the lepton momentum spectrum significantly as can be seen in Figures K. 23 - K. 24. ..... 180
7.11 Lepton momentum in the sideband region, here it is indicated that the MC is doing much better at simulating the background lepton momentum distribution. ..... 180
8.1 Plots for $\cos \theta_{\pi}$ (angle between the slow pion and $D^{*}$ direction) for data (points) and MC (histogram) in bins of $D^{*}$ center-of-mass momentum. In units of $\mathrm{GeV} / c$ the bins are (1) 0.5-1.0, (2) 1.0-1.5, (3) 1.5-2.0, (4) 2.0-2.5, (5) 2.5-3.0, (6) 3.0-3.5 and (7) 3.5-4.0. $B^{\prime}$ s can only produce $D^{*}$ momenta has high as bin 4 . Bins 1-3 are most relevant to this analysis. ..... 208
8.2 Distribution of $\delta m$ for MC (L) and Data (R). ..... 212
8.3 Ratio of Data to MC for the four kinematic variables in the high sideband region in Release 10 data, and best polynomial fits (varying in order from 4 to 6 depending on what best fitted the distribution). ..... 220
C. 1 Overlaying of generated MC (EvtGen, points) and fitting PDF (RooFit, curve), both with LLL parameters. ..... 239
C. 2 Overlaying generated MC (EvtGen, red crosses) and fitter MC (RooFit Toy MC, black histograms with error bars ) both with LLL parameters (note that the RooFit "curve" in this case and for all Figures C.2-C. 6 have been generated by throwing events with the theoretical PDF and binning them, not by taking a projection of the PDF itself as in Figure C.1, which is why the "curves" appear jagged).240

C. 3 Overlaying of generated MC (EvtGen, crosses) and fitter MC (RooFit Toy
MC, histograms with error bars ) both with LLL parameters; EvtGen has
FSR here (only in this case). ..... 242
C. 4 Overlaying of generated MC (EvtGen, crosses) and fitter MC (RooFit Toy MC, histograms with error bars ), both with CCC parameters. ..... 243
C. 5 Overlaying of LLL generated MC (EvtGen, crosses) and CCC fitter MC (RooFit Toy MC, histograms with error bars ). ..... 244
C. 6 Overlaying of generated (EvtGen, histogram) and fitting (RooFit, curve) PDF's - both with LLL parameters. ..... 245
F. $1 \Upsilon(4 S)$-based Kinematic Variable resolutions are in black, Y-frame resolu- tions in blue (dashed), top: left- $w$; right- $\cos \theta_{\ell} ;$ bottom: left- $\cos \theta_{V} ;$ right- $\chi$. In all cases the Y-frame resolutions are better, for $\cos \theta_{\ell}$ and $\cos \theta_{V}$ dramatically so (note that because the samples had the same number of events, all blue curves have been scaled down to fit within the frame). ..... 259
K. 1 Kinematic Variables in the signal window for electrons only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied. ..... 281
K. 2 Kinematic Variables in the sideband region for electrons only, these indi- cate the amount of agreement in the combinatorial background between Data and MC. ..... 281
K. 3 Kinematic Variables in the signal window for muons only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied. ..... 282
K. 4 Kinematic Variables in the sideband region for muons only, these indicate the amount of agreement in the combinatorial background between Data and MC. ..... 282
K. 5 Kinematic Variables in the signal window for the $D^{0} \rightarrow K^{-} \pi^{+}$mode only, these indicate decent agreement between both the true signal and back- grounds after all final cuts are applied. ..... 283
K. 6 Kinematic Variables in the sideband region for the $D^{0} \rightarrow K^{-} \pi^{+}$mode only, these indicate the amount of agreement in the combinatorial background between Data and MC. ..... 283
K. 7 Kinematic Variables in the signal window for the $D^{0} \rightarrow K^{-} \pi^{+} \pi^{+} \pi^{-}$mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied. ..... 284
K. 8 Kinematic Variables in the sideband region for the $D^{0} \rightarrow K^{-} \pi^{+} \pi^{+} \pi^{-}$ mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.
K. 9 Kinematic Variables in the signal window for the $D^{0} \rightarrow K \pi \pi^{0}$ mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.
K. 10 Kinematic Variables in the sideband region for the $D^{0} \rightarrow K \pi \pi^{0}$ mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.
K. 11 Kinematic Variables in the signal window for the $D^{0} \rightarrow K_{S} \pi \pi$ mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.286
K. 12 Kinematic Variables in the sideband region for the $D^{0} \rightarrow K_{S} \pi \pi$ mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.
K. 13 The distributions of $\cos \theta_{V}$ and $\cos \theta_{\ell}$ for different regions of $w$ in the signal window for both Data and MC after the final cuts are applied. Left-hand column is $\cos \theta_{V}$, right-hand column $\cos \theta_{\ell}$. Top row: upper half of $w$ range, bottom row,lower half of $w$ range.287
K. 14 The distribution of $\chi$ for different regions of $\cos \theta_{V}$ in the signal window for both Data and MC after the final cuts are applied. Left column top row is lowest quarter of $\cos \theta_{V}$ range, right column top row is the middle two quarters of $\cos \theta_{V}$ range and left column bottom row is the highest quarter of $\cos \theta_{V}$ range. Note that these distributions are also visible in the 2 dimensional correlation plots for $\cos \theta_{V}$ and $\chi$. Carefully studying the PDF will show that these plots show behavior concordant with it.
K. 15 Two dimensional correlation plots, MC are on the left, data on the right for the kinematic variables in the signal window. The top three pairs are between w and the three angles, the bottom three pairs between the three angles themselves. We see a good agreement between Data and MC here (since we do not overlay the plots, these and all following 2D plots are not normalized to the same number of events).
K. 16 Two dimensional correlation plots, MC are on the left, data on the right for the kinematic variables in the high sideband region. The top three pairs are between $w$ and the three angles, the bottom three pairs between the three angles themselves. We see a good agreement between Data and MC here, indicating good combinatorial background modeling in the MC
K. 17 The 2D correlation plots between the Kinematic Variables, shown is the Data histogram divided by the MC histogram for the signal window. The numbers given are the bin entries.

## xiv

$$
\begin{aligned}
& \text { K. } 18 \text { The 2D correlation plots between the Kinematic Variables, shown is the } \chi 2 \\
& \text { contribution per bin for the Data histogram divided by the MC histogram } \\
& \text { for the signal window. . . . . . . . . . . . . . . . . . . . . . . . . . . . . } 292
\end{aligned}
$$

K. 19 Slow pion $p_{T}$ for EvtGen produced files, the legend is as described in Sec- tion 7.2.1. Left are the XLL plots, Right are LXL. ..... 293
K. 20 These are the LLX plots for slow pion $p_{T}$ for EvtGen files ..... 293
K. 21 Slow pion momentum in the signal window, indicates our MC is doing a decent job at simulating the signal slow pion momentum distribution (with some of the variation coming from the form factor differences in data and MC) ..... 294
K. 22 Slow pion momentum in the sideband region, indicates our MC is doing well at simulating the background slow pion momentum distribution. ..... 294
K. 23 Lepton lab momentum for EvtGen produced files, the legend is as shown and described in Section 7.2.1. Left are the XLL plots, Right are LXL. ..... 295
K. 24 These are the LLX. ..... 295
K. 25 Comparison of kinematic variables of signal-type events in data and MC (all signal cuts as listed in Ch. 5). ..... 296
K. 26 Comparison of kinematic variables of $B^{ \pm}$-type background in data and MC (signal cuts except for $\cos \theta_{B Y}$ outside of the $[-1 \rightarrow 1]$ window). ..... 296
K. 27 Comparison of kinematic variables of fake lepton-type background in data and MC (all signal cuts except for non-VT electron or muon selected at final step). ..... 297
K. 28 Comparison of kinematic variables of uncorrelated lepton ( $D^{*}$ and leptonfrom different B's) type background in data and MC (signal cuts exceptfor $\cos \theta_{D^{*} \ell}>0$ selected, so $D^{*}$ and lepton point into same hemisphere,which is strongly disfavored in true $D^{*} \ell \nu$ decay)298
K. 29 Comparison of kinematic variables of combinatoric sideband region typebackground in data and MC (signal cuts except for high mass sidebandregion $0.165>\delta m>0.150 \mathrm{GeV}$ selected).298
L. 1 Fit distribution using 90 EvtGen calibration files, fitting to the same sample. 302
L. 2 Fit distribution using 100 pairs of background control files, fitting to the same initial sample generated with CLEO parameters (see text). ..... 308
N. 1 Distribution of fits with FSR resummed photon calibration files, fit to afile with FSR on - the distributions are not very Gaussian (this is only 10points), but what is important is the small size of the RMS spread.316

## List of Tables

1.1 Characteristics of fermions in the Standard Model broken down by first, second and third generations [1]. ..... 4
1.2 Characteristics of Force carrying Bosons in the Standard Model [1]. ..... 5
1.3 Partial list of baryons in ground state $\mathrm{SU}(4)$ multiplet [1]. The first row contains uncharmed baryons, the second row lists baryons with one charm quark, the third row lists baryons with two charm quarks while the last row contains baryons with three charm quarks. ..... 9
1.4 Pseudoscalar and vector mesons [1]. ..... 10
1.5 Particle masses and lifetimes especially useful for this analysis [1]. The daughters column shows either the primary decay or a decay which is used in this work. ..... 13
3.1 Ranges of $w, q^{2}$, and $\xi(w)$. ..... 30
4.1 SVT sublayers, radius and pitch (distance between each of the active sili- con detection strips in the given sublayer) ..... 52
4.2 DCH superlayers, layer type is axial(A) or stereo(U,V). ..... 54
4.3 Trigger $\phi$ map objects defined for the EMT and DCT. The IFT has a single $\phi$ map labelled U which indicates a track in the IFR (SL = DCH SuperLayer). 63
4.4 Trigger objects defined for the EMT and DCT. The IFT has a single object labelled U which indicates one or more tracks in the IFR. ..... 67
5.1 Composite particles used in this analysis, provided by CompositionTools. ..... 79
5.2 Electron particle ID category requirements (the $\Delta \phi$ and DIRC columns in- dicate whether these variables are used for the given category). ..... 86
5.3 Muon particle ID category requirements. ..... 88
5.4 Summary of $D^{0}$ cuts. ..... 94
5.5 Summary of $D^{*}$ and $D^{*} \ell \nu$ cuts. ..... 107
5.6 Branching fractions used by EvtGen (release 8.8.0c-physics-1). ..... 115
5.7 Calculating the product of detector acceptance, track efficiency and Com- position Tools cuts efficiencies, by decay mode ..... 115
5.8 Summary of no cut, $D^{0}$ cut, $D^{*}$ cut and $D^{*} \ell \nu$ cut efficiency. This table list the number of signal and background events remaining after the associ- ated cuts have been applied. Note that there is no $D^{*}-D^{0}$ cut applied (see next table for that cut). ..... 116
5.9 Summary of no cut, $D^{0}$ cut, $D^{*}$ cut and $D^{*} \ell \nu$ cut efficiency. This table list the number of signal and background events remaining after the associ- ated cuts have been applied. Note that there is no $D^{*}-D^{0}$ cut applied (see next table for that cut). ..... 116
6.1 Calibration on signal SP4Tot, Fit to SP4A. ..... 148
6.2 Calibration on signal SP4Tot, Fit to SP4B. ..... 148
6.3 Calibration on signal SP4Tot, Fit to SP4C. ..... 149
6.4 Calibration on signal SP4Tot, Fits, from left to right, of: first row: A-B, A-C, second row: B-A, B-C, third row: C-A, C-B (all in form: [ SP4X Reco cut - (scale factor) * SP4Y MC truth background], where $\mathrm{X}, \mathrm{Y}=$ one of $\mathrm{A}, \mathrm{B}, \mathrm{C}$ ). Given are the input values, the output fit values, and the statistical error as returned by Minuit. ..... 157
$6.5 \chi^{2}$ and $\chi^{2}$-probability for kinematic variable $1 D$ projections and lepton momentum; the number of bins in these plots is 10 . The number of degrees- of-freedom is 9 , since we have forced the normalization to be equal be- tween data and MC. ..... 162
7.1 Ratios of number of event yields in Release 10 data over those in SP4 MC, in signal window and sideband regions. ..... 171
8.1 Varying the parameters used in calibration file and fitting to the file with all the input R parameters set to 1.000 (EvtGen). The numbers in the topmost row refer to the values of the parameters used to generate the calibration file (the same for all three parameters $R_{1}, R_{2}, \rho^{2}$, here). All files are on the order of 100 k events, no cuts, and the quoted Minuit errors are $\sim .015$ for each parameter ..... 188
8.2 Results of fits to MC and Data $\delta m$ distributions. ..... 212
8.3 All the semileptonic $B$ decay modes. The first column is the BR (in per- cent) put into the MC when this mode is generated from generic $B$ decays, the PDG (2003) measurement is given if available, and the third column is the percentage variation used. The last column is the error per mode computed from the variations. Modes related by isospin between $B^{0}$ and $B^{+}$decays have been treated together. ..... 218
8.4 Changes in the fitted parameters for reweighting of the MC combinatoric background distributions in the four kinematic variables, as shown in Fig- ure 8.3. ..... 221
8.5 Systematic errors estimates table (errors contributing negligible amounts: $D^{0}$ tracking and DeltaM MC/Data differences, are not shown in this table; see text) ..... 222
8.6 MC statistical errors for our $\sim 200 \mathrm{fb}^{-1}$ equivalent SP4 generic MC sample. ..... 222
8.7 MC statistical and normal statistical errors per unit $\mathrm{fb}^{-1}$ of MC luminos- ityfor the MC statistical error, or per 1 k events in the case of the actual statistical error (luminosity cannot be directly compared between CLEO and Babar because CLEO only reconstructed two $D^{0}$ decay modes, where Babar reconstructs four, but for Babar roughly 1 k signal events are recon- structed per $\mathrm{fb}^{-1}$ ). At a naive level, statistical errors at a collected data luminosity $\mathcal{L}_{2}$ can be scaled from the above SP4 statistical errors simply by multiplying by the factor $\sqrt{\frac{1}{\mathcal{L}_{2}}}$, while MC statistical errors at a MC equiv- alent luminosity of $\mathcal{L}_{M C}$ can be scaled from the above SP4 MC statistical errors by multiplying by the factor $\sqrt{\frac{1}{\mathcal{L}_{M C}}}$. . . . . . . . . . . . . . . . . 223
C. 1 LLL file, 100k events - see 7.2.1 for parameter labelling nomenclature ..... 241
C. 2 AAA file, 100k events ..... 241
C. 3 EEE file, 100k events ..... 241
C. 4 CCC file, 10 k events ..... 241
C. 5 CCC file, 100k events ..... 244
C. 6 CCC file, 1 M events ..... 245
D. 1 Calibrating on one EvtGen sample, then fitting to five statistically inde-pendent EvtGen samples with no cuts. Inputs are $R_{1}=R_{2}=\rho^{2}=1.00 . .247$
D. 2 Analytic calibration, fit to 111 k file with cuts. ..... 248
D. 3 Moment method calibration from a large (1M) file with cuts in it, fit to same 111k file with cuts. ..... 248
D. 4 Calibrating on one EvtGen sample, then fitting to five statistically inde-pendent EvtGen samples with cuts as specified in the text. Inputs are$R_{1}=R_{2}=\rho^{2}=1.00$.249D. 5 Calibrating on one EvtGen sample of 1 M events generated from input val-ues $R_{1}=R_{2}=\rho^{2}=1.00$, then fitting to four statistically independentEvtGen samples with varying parameters, and cuts as specified in the text.Minuit statistical errors shown with the fit results250
D. 6 Calibrating on multiple EvtGen samples of 1M events generated with in-put parameter values as shown, then fitting to an EvtGen sample withinput values $R_{1}=R_{2}=\rho^{2}=1.00$. All files have cuts as specified in thetext. Minuit errors shown with the fit results.250
E. 1 Calibration on various lepton files: Fits on left are to electron-only file, on right to muon only file. Calibration from top to bottom: on file with both leptons, on electron-only file, and on a muon only file (note the fit and calibration files are only on partially different statistically independent files in this case). The errors are the Minuit-quoted statistical errors only, for this and all following tables in this Appendix.
E. 2 All fits to mode $1(K \pi)$, calibration on various $D^{0}$ decay mode files. Calibration from top to bottom: all modes combined, mode 1,2 , and $3(1=K \pi$, $2=K 3 \pi, 3=K \pi \pi^{0}$ ) (note the fit and calibration files are only on partially different statistically independent files in this case).
E. 3 All fits to mode $2(K 3 \pi)$,calibration on various $D^{0}$ decay mode files. Calibration from top to bottom: all modes combined, mode 1,2 , and $3(1=K \pi$, $2=K 3 \pi, 3=K \pi \pi^{0}$ ). (note the fit and calibration files are only on partially different statistically independent files in this case).
E. 4 All fits to mode $3\left(K \pi \pi^{0}\right)$,calibration on various $D^{0}$ decay mode files . Calibration from top to bottom: all modes combined, mode 1,2 , and $3(1=K \pi$, $2=K 3 \pi, 3=K \pi \pi^{0}$ ). (note the fit and calibration files are only on partially different statistically independent files in this case).256

H. 1 Calibration on Evtgen sample with all $\mathrm{R}=1.00$, fit to statistically indepen
dent sample of the same type. ..... 263
H. 2 Calibration on Evtgen sample with all $\mathrm{R}=1.00$, fit to statistically independent sample of the same type with 20 k of added events with all $\mathrm{R}=0.50$, fit is very off as expected.

$$
264
$$

H. 3 Calibration on Evtgen sample with all $\mathrm{R}=1.00$, fit to statistically independent sample of the same type with 20 k of added events with all $\mathrm{R}=0.50$, and 20 k statistically independent events with all $\mathrm{R}=0.50$ also subtracted off - fit is spot on.

$$
264
$$

J. 1 The $\frac{\chi^{2}}{\text { dof }}$ for EvtGen samples, the first two columns are the samples compared, and the third is the $\frac{\chi^{2}}{\text { dof }}$. Sample description: Rlll.0, 2,6 are 112k event statistically independent samples generated with the CLEO parameters (as per the Standard Form Factor Parameter Notation of Section 7.2.1), and with cuts applied. Rlll. $2^{\prime}, 4^{\prime}$ are 100 k event statistically independent samples generated with the CLEO parameters and no cuts. Raaa,bbb,ccc,ddd,eee,lll" are 100k event samples generated with various sets of parameters. See text for more details.
J. 2 The $\frac{\chi^{2}}{\text { dof }}$ for SP4 samples, the first two columns are the samples compared, and the third is the $\frac{\chi^{2}}{\text { dof }}$. Naming scheme for samples: SP4B.PK is the signal region piece of SP4B (which has both signal and residual background), SP4B.SIG is the MC-truth selected signal in the signal region piece of SP4B, SP4B.PKBKGD is the MC-truth selected peaking background piece of SP4B, and SP4B.SB is the sideband region sample from SP4B, and the same nomenclature holds for SP4A.277
J. 3 The $\frac{\chi^{2}}{\text { dof }}$ for background control samples in data vs. MC. See text in Section J. 3 for more details.
N. 1 Analytic function calibration on LLL Evtgen sample with FSR off. . . . . . 312
N. 2 Analytic function calibration on LLL Evtgen sample with FSR on. . . . . . 312
N. 3 Real calibration on another LLL Evtgen sample (no FSR) with cuts (see text). 313
N. 4 Real calibration on another LLL Evtgen sample (FSR on) with cuts (see text). 313
N. 5 Real calibration on another LLL Evtgen sample (FSR on) with cuts (see text). 314
N. 6 Real resummed calibration on another LLL Evtgen sample with cuts (see text). 314

## Acknowledgements

I want to thank the following people for extensive help and advice provided:
My advisor: R.Jacobsen
Other members of my qualifying exam and thesis committees: N. Arkani-Hamed, R.Cahn, E.Haller, N.Roe, M.Shapiro.

My primary collaborators, so very essential in so many ways to the successful completion of this work: A.Snyder, K.Tackmann.

My other thesis readers get special mention for improving the text and structure of this work: M.C.Simani, M.Mazur, S.Kaur, K.Mishra.

Analysis Working Group and Review Committee members: D.Bernard, G.Blaylock, L.Wilden, S.Mazzoni; F.Simonetto, R.Kowaleski..

Theorists whose work is directly related to our results and with whom we had helpful discussions: Z.Ligeti, M.Neubert.

Other BaBarians, many of whom have been instrumental in achieving this end, and all of whom have been helpful: J.Albert, A.Borgland, B.Brau, A.Breon, D.Brown, M.Bruinsma, P.Burchat, E.Charles, C.Cheng, K.Copic, J.Dingfelder, V.Erkcan, R.Faccini, S.Gowdy, D.Kirkby, N.Kuznetsova, C.LeClerc, R.Liu, V.Luth, G.F.F.Mancinelli, T.Meyer, A.Mihalyi, J.Perl, M.Pripstein, S.Rahatlou, J.Richman, C.Roat, A.Ryd, K.Schubert, S.Sekula, S.Sen, F.Simonetto, A.Soffer, M.Sokoloff, A.Telnov, A.Weinstein.

Other physicists who have been important along the way: S.Brodsky, H. Steiner, D. Loomba, and members of the LBL and SLAC theory groups.

Last, I also need to express my profoundest gratitude to others without whom the travels along this Path, and who I am today, would not have been the same, and who were thus integral to this work as well:

First, and foremost - to my Mom, Dad, and Sister.
Then next to all those in my sprawling family (and near-family): all the Aujlas, Beasleys, Dhesis, Gills, Khairas, Mahangers, Mahavnis, Sandhus, Shergills, Sohis, Thinds, and to all others spread across (at least) three continents.

And further to (not in any strict order): N.Dhesi, B.Rittenhouse, L.J.H.Heydemann, L.M.Dietrich, N.Sandhu, S.Stansbury, C.Savage-Sachs, T.Ciszewski, M.Singh, K.Kang, P.Randhawa, H.Gundling, A.Hurley, J.Pehling, R.Dave, J.Holbrook, S.Meghrouni, V.Chalcraft, L.Carswell, D.Pierce, M.Riordan, J.Bhuller, S.Riordan, S.O'Riordan. P.Anderson, R.Anderson, S.Schulten, C.Farrell, E.Reyes, D. and L.Cooke and all other CVians. T.Miller, R.Todaro and other fine SLAC-folk. C.King, G.Weston, D.Serres, N.Robertson, K.Lebacqz, E.Goodwin, M.Lappe, E.Leslie, F.Fleischman, J.Gerner, J.Pratt, N.Matthews, Mrpl.Roberts, and the rest of Los Ratones. S.Leitman, A.Hernandez, L.McVittie, C.Balz, M.Mohanan, D.Solnit, S.Khan, S.Buttar, V.Brar, A. Maloney, L.Kemmerer, and the other good folks of BSAL, SCPJ, DASW, EF, O2C, T.R., S.Fire, UnBC, F.Tower, BoR, UJB, ToG, M.Dew, and several generations of Baens, Synners, and Thetans Chi. And to everyone I have not mentioned specifically by name - you know who you are, and how each of you has left your mark on my livingway. With deep gratitude, always - Thank You.

Path. : - >

## Chapter 1

## Introduction and Motivations

"We are a way for the universe to know itself. Some part of our being knows this is where we came from. We long to return. And we can, because the Cosmos is also within us. We're made of star stuff." - Carl Sagan ${ }^{1}$
"There is nothing new to be discovered in physics now. All that remains is more and more precise measurement." - Lord Kelvin [2] ${ }^{2}$

These two quotes come from two competing and rather differing points of view that natural scientists may hold: the former an exploratory, adventurous and somewhat humble one, and the latter a rather mechanistic and fairly arrogant one. The particle physicist of today most likely comes now from a very different worldview than Lord Kelvin, and probably one much closer to that of Sagan.

Unlike Kelvin who asserted that there were no major advances left to make in

[^0]physics, many particle physicists today actively seek to unseat the currently accepted and spectacularly successful "Standard Model of Particle Physics" ${ }^{3}$. There is great anticipation for the next generation of accelerators (the accelerators at Fermilab near Chicago, Illinois and at CERN ${ }^{4}$ in Geneva, Switzerland) to possibly discover "physics beyond the Standard Model". Even current experiments like Belle and BaBar have the potential of uncovering something new.

### 1.1 Standard Model

The Standard Model has great computational powers. It can be used to calculate particle lifetimes, reaction rates, cross sections, etc. and calculations using it can be done with tremendous accuracy. The weaknesses of the Standard Model include the fact that the Standard Model neglects gravity and also that it critically depends on a set of 18 parameters [6], which all must be measured, not derived from basic principles. These 18 parameters include the six quark masses, three lepton masses, four elements from quark mixing (Cabibbo-Kobayashi-Maskawa (or CKM) matrix [13] entries), two parameters involving the yet-to-be discovered Higgs particle (the Higgs mass and vacuum expectation value) and three parameters which describe the coupling strengths of the strong, weak and electromagnetic forces - the three fundamental forces that are important in the realm of particle physics (the fourth fundamental force, gravitation, is, as we will see, too weak

[^1]to have any impact in particle physics).
The direct measurement of neutrino mass (which appears imminent in the next few years) will inevitably expand this number of parameters by at least seven (three neutrino masses and four mass mixing parameters - six mixing parameters if the neutrino is a Majorana particle). This plethora of "fundamental numbers" leads physicists to believe that the Standard Model is an approximation of a larger, more encompassing theory, one that likely includes gravity and has only a few parameters.

Much of the theoretical framework of Standard Model was already in place during the 1970's. The confirmation of the Standard Model with remarkable precision occurred during the 1980's and 1990's [7]. The next several sections of this chapter are devoted to describing the Standard Model, as a prelude to the parameters of the Standard Model we will be focussing on in the rest of this work.

### 1.1.1 Elementary Particles - quarks, leptons and forces

The Standard Model attempts to describe all the particles and interactions which make up our universe. The particles and their interactions are concisely described by listing the particle's quantum numbers - fundamental attributes intrinsic to elementary particles. The relevant quantum numbers include the particle's mass, charge, "color charge", spin and flavor. All fundamental particle masses (except the ones identically equal to zero) are given by measuring 11 of the 18 free parameters of the Standard Model. The charge describes a particle's coupling with the electromagnetic force while flavor is im-
portant for the weak force. Color charge determines the strong interactions and the spin describes the type of statistics the particle will obey [8].

From quantum mechanics, all particles can be broken into two classes based on their spin. Particles with half-integer units of spin are called fermions and they obey Fermi statistics. The most important and familiar result from Fermi statistics is the Pauli Exclusion Principle. This principle states that no two identical fermions can occupy the same state at the same time, i.e., have the exact same quantum numbers. Particle with integer spins are called bosons. Not only can two identical bosons occupy the same state at the same time, multiply occupied states are favored over singly occupied states.

Table 1.1 shows leptons and quarks that make up matter. All the quarks and leptons are fermions. All the force-carrying particles in Table 1.2 are bosons.

| Leptons (spin=1/2 $)$ |  |  | Quarks (spin=1/2 $\hbar)$ |  |  |
| :--- | ---: | ---: | :--- | ---: | ---: |
| Flavor | Mass $\left(\mathrm{GeV} / c^{2}\right)$ | Charge | Flavor | Mass $\left(\mathrm{GeV} / c^{2}\right)$ | Charge |
| $\nu_{e}$ | $<3 \times 10^{-9}$ | 0 | u up | $0.001-0.005$ | $2 / 3$ |
| e | 0.000511 | -1 | d down | $0.003-0.009$ | $-1 / 3$ |
| $\nu_{\mu}$ | $<2 \times 10^{-4}$ | 0 | c charm | $1.15-1.35$ | $2 / 3$ |
| $\mu$ | 0.106 | -1 | s strange | $0.075-0.170$ | $-1 / 3$ |
| $\nu_{\tau}$ | $<2 \times 10^{-2}$ | 0 | t top | $174.3 \pm 5.1$ | $2 / 3$ |
| $\tau$ | 1.777 | -1 | b bottom | $4.0-4.4$ | $-1 / 3$ |

Table 1.1: Characteristics of fermions in the Standard Model broken down by first, second and third generations [1].

There are three generations or families of particles. Each generation has two quarks, one with charge $+2 / 3$ (an "up" type quark) and one with charge $-1 / 3$ ("down" type quark), a negatively charged lepton and a chargeless neutrino. Each member of the family has an antimatter partner that has the same spin and mass, but opposite charge

| Interaction | Gravity | Weak | Electromagnetic | Strong |
| :--- | :---: | :---: | :---: | :---: |
| Acts on | Mass-Energy | Flavor | Electric Charge | Color Charge |
| Particles | All | Quarks, | Electrically | Quarks, |
|  |  | Leptons | Charged | Gluons |
| Force carrying | Graviton | $W^{+}, W^{-}, Z^{0}$ | $\gamma$ | Gluons |
| Boson |  |  |  | $\left(g_{1}, \ldots, g_{8}\right)$ |
| Spin | $2 \hbar$ | $\hbar$ | $\hbar$ | $\hbar$ |
| Charge | 0 | $\pm 1,0$ | 0 | 0 |
| Mass $\left(G e V / c^{2}\right)$ | 0 | $\pm 80.4,91.2$ | 0 | 0 |

Table 1.2: Characteristics of Force carrying Bosons in the Standard Model [1].
and flavor quantum numbers. The family members get progressively heavier as one proceeds from the first to second to third generation. These particles are currently considered fundamental, although it is possible they are made up of smaller entities that only reveal themselves at extraordinarily high energies.

The four forces of nature all have associated particles which mediate all interactions. The graviton is a spin-2 particle which has yet to be detected (the only other undetected particle accepted as a normal part of the Standard Model is the Higgs Boson). Only massless particles produce long-range inverse square force laws. The graviton is assumed to be massless because of the long range of gravity. The long-range nature of electromagnetism implies that the photon is also massless (which in this case also has direct experimental confirmation [1]). However, there are several striking differences between electromagnetism and gravity. One is is that gravity is purely attractive while electromagnetism is attractive and repulsive. A second difference is their relative strengths. If we compare their dimensionless coupling constants, $G_{N} M^{2} / \hbar c \approx 6.7 \times 10^{-39}\left(\mathrm{GeV}^{2}\right) \times M^{2}$
for gravity and $\alpha=e^{2} / \hbar c \approx 1 / 137$ for electromagnetism [1], we find that gravitational strength would be roughly equal to electromagnetic strength for a particle with unit charge and a mass $M \approx 10^{18} \mathrm{GeV} / \mathrm{c}^{2}$. Yet today's largest particle accelerators reach interaction energies $\sim 10^{3} \mathrm{GeV}$, and the Large Hadron Collider (LHC) will reach $\sim 10^{4} \mathrm{GeV}$ around 2007. The gravitational attraction between two protons is $\sim 10^{36}$ times smaller than the electromagnetic repulsion. The disparity is much larger for lighter particles like the electron or light quarks. Thus, this vast difference in relative strength makes gravity all but irrelevant on the subatomic level until one reaches energies approaching the Planck scale (the energy where all forces are expected to be equally important $\approx 10^{19}$ GeV ).

The weak force is mediated by the $W$ and $Z$ bosons. The large masses associated with these particles means that the force can only act over a short range, roughly estimated by the inverse of the force carrying particle's mass. The weak force and electromagnetism can be combined into a single electroweak force which contains four gauge bosons and is described by the $\mathrm{SU}(2) \times \mathrm{U}(1)$ symmetry group. The coupling strengths for both the weak force and the electromagnetic force are significantly smaller than unity, which allows for perturbative calculations [8]. The electroweak force will be discussed further in Section 1.1.5.

The strong force is described by Quantum Chromodynamics with the $\mathrm{SU}(3)$ symmetry group. Each quark has a "color charge" with one of three values, red(r), green $(\mathrm{g})$, or blue $(\mathrm{b})$. An antiquark carries anticolor, $\bar{r}, \bar{g}$, or $\bar{b}$. Gluons carry and trans-
fer color charge. Only quarks and gluons carry color, so only these particles participate in strong interactions. At high energies (e.g., at the $Z$ mass of $91 \mathrm{GeV} / \mathrm{c}^{2}$ ), the strong coupling constant is relatively small (0.1) compared to unity and perturbative calculations can be done. However, at lower energies the coupling constant "runs" to larger values making secondary interactions more important thus invalidating any perturbative calculations. One consequence of the large value of the strong coupling constant at low interaction energies (i.e., anything far below the $Z$ pole mass at $91 \mathrm{GeV} / c^{2}$ ) is the non-perturbativity of the theory, which leads to the inability to do calculations of any processes using the full machinery of QCD. Instead, simpler models or approximations of the internal wavefunction structure of these particles and the interactions between them must be used to make predictions, and one of these models is Heavy Quark Effective Theory, which we will learn more about in Section 1.2.1.

Another consequence of large $\alpha_{s}$ at low energies is "color confinement", which states that bound particles can only be colorless (color singlet) so no free quarks or gluon can be found. Particles can be colorless if they are composed of three quarks of each of the three color charges (r,g,b) or of a quark and antiquark carrying opposite color charge (e.g. red $+\overline{r e d}$ ). When quarks in a baryon or meson are separated, during a collision in an accelerator for example, it is energetically favorable for a quark-antiquark pair to appear from the vacuum and bind up with the bare quarks to make colorless final-state observable objects. So far, quarks have only been found as baryons or mesons [8] (though recent evidence is pointing to the transient existence of (still colorless) 5-quark bound
states). We now turn to describing briefly the properties of these bound-state particles.

### 1.1.2 Composite QCD Particles - Baryons and Mesons

A "hadron" is any particle composed of quarks, which breaks down further into the two categories of "baryons" and "mesons".

### 1.1.2.1 Baryons

A baryon is a three quark color singlet state (rgb) and a meson is a quarkantiquark state (color $-\overline{\text { color }}$ ). There are six different quark flavors listed in table 1.1. If one considers $n$ different flavors, there exists an approximate $S U(n)$ flavor symmetry. This symmetry is broken by the different quark masses. The first two quarks have very similar masses, and the $\mathrm{SU}(2)_{\text {flavor }}$ symmetry (isospin) is a very good symmetry as shown by the small mass difference between protons and neutrons. The strange quark is modestly heavier than the up and down quarks, giving an approximate $\mathrm{SU}(3)_{\text {flavor }}$ symmetry for ( $\mathrm{u}, \mathrm{d}, \mathrm{s}$ ) quarks and all higher $\mathrm{SU}(\mathrm{n})_{\text {flavor }}$ symmetries are badly broken by the heavy quarks.

Baryons are fermions, which by definition have an antisymmetric wave function. The $\mathrm{SU}(3)$ color singlet state is a completely antisymmetric state, so a baryon wave function can be decomposed as follows:

$$
\begin{equation*}
|q q q>=| \text { color singlet }>_{A} \times \mid \text { spin, flavor, space }>_{S} \tag{1.1}
\end{equation*}
$$

The color quantum number determines the types of baryons that may be formed.

Without the antisymmetric color component, the $\Delta^{++}$and other baryons with three identical quarks would be impossible to form since the space, spin and flavor part of the wave function would be symmetric [8].

| $\mathrm{SU}(3)$ | quark content | $\mathrm{SU}(3)$ | quark content |
| :---: | :---: | :---: | :---: |
| $\mathrm{n}, \mathrm{p}$ | $\mathrm{udd}, \mathrm{uud}$ | $\Delta^{-}, \Delta^{0}, \Delta^{+}, \Delta^{++}$ | ddd, udd, uud,uuu |
| $\Sigma^{-},\left(\Sigma^{0}, \Lambda^{0}\right), \Sigma^{+}$ | $(\mathrm{dd}, \mathrm{ud}, \mathrm{uu}) \times \mathrm{s}$ | $\Sigma^{-}, \Sigma^{0}, \Sigma^{+}$ | $(\mathrm{dd}, \mathrm{ud}, \mathrm{uu}) \times \mathrm{s}$ |
| $\Xi^{-}, \Xi^{0}$ | $(\mathrm{~d}, \mathrm{u}) \times \mathrm{ss}$ | $\Xi^{-}, \Xi^{0}$ | $(\mathrm{~d}, \mathrm{u}) \times \mathrm{ss}$ |
|  |  | $\Omega^{-}$ | sss |
| $\Sigma_{c}^{0},\left(\Sigma_{c}^{+}, \Lambda_{c}^{+}\right), \Sigma_{c}^{++}$ | $(\mathrm{dd}, \mathrm{ud}, \mathrm{uu}) \times \mathrm{c}$ | $\Sigma_{c}^{0}, \Sigma_{c}^{+}, \Sigma_{c}^{++}$ | $(\mathrm{dd}, \mathrm{ud}, \mathrm{uu}) \times \mathrm{c}$ |
| $\Xi_{c}^{0}(2), \Xi_{c}^{+}(2)$ | $(\mathrm{d}, \mathrm{u}) \times \mathrm{sc}$ | $\Xi_{c}^{0}, \Xi_{c}^{+}$ | $(\mathrm{d}, \mathrm{u}) \times \mathrm{sc}$ |
| $\Omega_{c}^{0}$ | ssc | $\Omega_{c}^{0}$ | ssc |
| $\Xi_{c c}^{+}, \Xi_{c c}^{++}$ | $(\mathrm{d}, \mathrm{u}) \times \mathrm{cc}$ | $\Xi_{c c}^{+} \Xi_{c c}^{++}$ | $(\mathrm{d}, \mathrm{u}) \times \mathrm{cc}$ |
| $\Omega_{c c}^{+}$ | scc | $\Omega_{c c}^{+}$ | scc |
|  |  | $\Omega_{c c c}^{++}$ | ccc |

Table 1.3: Partial list of baryons in ground state $S U(4)$ multiplet [1]. The first row contains uncharmed baryons, the second row lists baryons with one charm quark, the third row lists baryons with two charm quarks while the last row contains baryons with three charm quarks.

The baryons that can be constructed from the lightest three quarks obey the approximate $\mathrm{SU}(3)_{\text {flavor }}$ symmetry and belong to multiplets described by:

$$
\begin{equation*}
3 \otimes 3 \otimes 3=10_{S} \oplus 8_{M} \oplus 8_{M} \oplus 1_{A} \tag{1.2}
\end{equation*}
$$

Where the S, M and A denote symmetric, mixed or antisymmetric states, under the interchange of any two quarks. Table 1.3 shows baryons extended to $\operatorname{SU}(4)$, which is an approximate flavor symmetry of any baryon composed of $u, d, s, c$ quarks. The first column is a 20-plet based on a $\mathrm{SU}(3)$ octet. The second column of baryons is based on the totally symmetric $S U(3)$ decuplet. A similar extension can be made to include bottom quarks in an approximate $\mathrm{SU}(5)$ flavor symmetry.

### 1.1.2.2 Mesons

The mesons are inherently simpler since they are composed only of two quarks and these are immediately not identical - one is a quark and the other necessarily an antiquark (in general the word "quark" is taken to be either, unless referred to in a context where it is contrasted with antiquarks specifically). Being made of just two quarks each with spin $1 / 2$, the meson can have a spin $S$ of 0 or 1 . For states in which the orbital angular momentum ( L ) is zero, all mesons will be either pseudoscalar particles, $J^{P}=0^{-}$, or vector particles, $J^{P}=1^{-}$. For the 3 lightest quarks, the multiplets are described by the color multiplets:

$$
\begin{equation*}
3 \otimes \overline{3}=8 \oplus 1 \tag{1.3}
\end{equation*}
$$

This tells us that both the pseudoscalar and vector mesons will form an octet plus singlet state. Again this can be extended for the charm and bottom quarks, as shown in the second and third boxrows in Table 1.4

|  | quark content | pseudoscalar mesons | vector mesons |
| :--- | :---: | :---: | :---: |
| $\mathrm{u}, \mathrm{d}, \mathrm{s}$ | $\mathrm{u} \bar{d},(\mathrm{u} \bar{u}-\mathrm{d} \bar{d}), \mathrm{d} \bar{u}$ | $\pi^{+}, \pi^{0}, \pi^{-}$ | $\rho^{+}, \rho^{0}, \rho^{-}$ |
| Mesons | $(\mathrm{u} \bar{u}+\mathrm{d} \bar{d}), \mathrm{s} \bar{s}$ | $\eta^{\prime}, \eta$ | $\omega, \phi$ |
|  | $\overline{\mathrm{u}}, \mathrm{s} \bar{u}, \bar{s} \mathrm{~d}, \mathrm{~s} \bar{d}$ | $K^{+}, K^{-}, K^{0}, \bar{K}^{0}$ | $K^{*+}, K^{*-}, K^{* 0}, \bar{K}^{* 0}$ |
| Charmed | $\mathrm{c} \bar{d}, \bar{c} \mathrm{~d}, \mathrm{c} \bar{u}, \overline{\mathrm{c}} \mathrm{u}$ | $D^{+}, D^{-}, D^{0}, \bar{D}^{0}$ | $D^{*+}, D^{*-}, D^{* 0}, \bar{D}^{* 0}$ |
| Mesons | $\mathrm{c} \bar{s}, \bar{c} \mathrm{~s}$ | $D_{s}^{+}, D_{s}^{-}$ | $D_{s}^{*+}, D_{s}^{*-}$ |
|  | $\mathrm{c} \bar{c}$ | $\eta_{c}$ | $\mathrm{~J} / \psi$ |
| Bottom | $\bar{b} \mathrm{u}, \mathrm{b} \bar{u}, \bar{b} \mathrm{~d}, \mathrm{~b} \bar{d}$ | $B^{+}, B^{-}, B^{0}, \bar{B}^{0}$ | $B^{*+}, B^{*-}, B^{* 0}, \bar{B}^{* 0}$ |
| Mesons | $\overline{\mathrm{s}}, \mathrm{b} \bar{s}$ | $B_{s}^{0}, \bar{B}_{s}^{0}$ | $B_{s}^{* 0}, \bar{B}_{s}^{* 0}$ |
|  | $\mathrm{~b} \bar{b}$ | $\eta_{b}$ | $\Upsilon(1 \mathrm{~S})$ |

Table 1.4: Pseudoscalar and vector mesons [1].

### 1.1.3 Particle decays

Most baryons and mesons decay into lighter elementary particles with relatively short lifetimes. Particles can decay via the strong, weak or electromagnetic forces. The characteristic lifetimes are generally determined by the force mediating the decay, but the kinematics of the decay and various symmetries can also greatly affect the result. The typical lifetime of a strong decay is $10^{-23} \mathrm{sec}$, electromagnetic is $10^{-16} \mathrm{sec}$ and weak decay is $10^{-8} \mathrm{sec}$. The only completely stable particles (as far as we currently know) are the proton, electron, photon and the neutrinos. All heavier baryons will ultimately decay into a proton plus extra stable particles (photons, leptons and neutrinos). All mesons eventually decay into leptons and neutrinos. The $\tau$ and $\mu$ leptons follow similar decay chains, always leaving an electron plus neutrinos and possibly photons.

The decay of any particle can be neatly broken into the product of two pieces, one representing the pure kinematics of the decay and the other being the matrix element between the initial and final states which carries the information about the physics involved in the particular decay. For the decay $X \rightarrow 1+2+\ldots+n:[7]$

$$
\begin{equation*}
d \Gamma=\frac{1}{2 E_{X}}|\mathcal{M}|^{2} \frac{d^{3} p_{1}}{(2 \pi)^{3} 2 E_{1}} \cdots \frac{d^{3} p_{n}}{(2 \pi)^{3} 2 E_{n}}(2 \pi)^{4} \delta^{(4)}\left(p_{X}-p_{1}-\ldots-p_{n}\right) \tag{1.4}
\end{equation*}
$$

Where $|\mathcal{M}|=<f|\mathcal{H}| i>$ is the matrix element. For the special case of a two body decay,

$$
\begin{equation*}
\Gamma(X \rightarrow 1+2)=\frac{p_{f}}{32 \pi^{2} m_{X}^{2}} \int|\mathcal{M}|^{2} d \Omega \tag{1.5}
\end{equation*}
$$

All angular dependence is carried by the matrix element $\mathcal{M}$. The final momentum of both daughters is equal in the rest frame of the particle $X$ and is denoted by $p_{f}$.

The total decay width of a particle is the sum of all individual channels, [7]

$$
\begin{equation*}
\Gamma_{t o t a l}=\sum \Gamma_{i} \tag{1.6}
\end{equation*}
$$

The decay rate is related to a particles lifetime through the equation

$$
\begin{equation*}
N_{X}(t)=N_{X}(0) e^{-\Gamma t}, \text { thus } \tau=1 / \Gamma \tag{1.7}
\end{equation*}
$$

Table 1.5 shows the mass, lifetime and decays of some particles which are particularly relevant to this analysis (which we will specify in more detail in Chapter 4). One interesting thing one can gather from the Table is which particles can be expected to be seen physically travelling from the decay vertex of a parent and which decay almost instantaneously. The BABAR detector tracking resolution is on the order of tens of microns (see Section 4.3). Thus we should clearly be able to see the $B^{0}$ and $D^{0}$ fly away from the parent and produce a separate decay vertex. The $D^{*}$ decays so rapidly that the daughters of the $D^{*}$ parent (a pion and $D$ meson) all appear to come from the same vertex.

### 1.1.4 CP violation

There are three discrete operators that are potential symmetries of the Standard Model Lagrangian. Two operators are space-time symmetries, Parity $(\mathrm{P})$ and Time rever$\operatorname{sal}(\mathrm{T})$. The Parity operation takes all space coordinates and replaces them with oppositely signed quantities $-(\mathrm{x}, \mathrm{t}) \rightarrow(-\mathrm{x}, \mathrm{t})$. This has the effect of reversing the handedness of a particle, similar to viewing the particle's mirror image. The time reversal operator replaces the time coordinate with an oppositely sign value $-(x, t) \rightarrow(x,-t)$, reversing the direction of

| meson | mass $(\mathrm{MeV})$ | lifetime (sec) | $\mathrm{c} \tau(\mu \mathrm{m})$ | Daughters | $\Gamma_{i} / \Gamma \%$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |
| $\pi^{ \pm}$ | 139.57 | $2.6 \times 10^{-8}$ | $7.8 \times 10^{6}$ | $\mu^{ \pm} \nu_{\mu}$ | 99.99 |
| $\pi^{0}$ | 134.98 | $8.4 \times 10^{-17}$ | 0.025 | $2 \gamma$ | 98.80 |
| $K^{ \pm}$ | 493.68 | $1.2 \times 10^{-8}$ | $3.7 \times 10^{6}$ | $\mu^{ \pm} \nu_{\mu}$ | 63.51 |
|  |  |  |  | $\pi^{ \pm} \pi^{0}$ | 21.16 |
| $K_{S}$ | 497.67 | $0.89 \times 10^{-10}$ | 2679 | $\pi^{+} \pi^{-}$ | 68.61 |
|  |  |  |  | $\pi^{0} \pi^{0}$ | 31.39 |
| $D^{0}$ | 1864.5 | $4.1 \times 10^{-13}$ | 123.7 | $K^{-} \pi^{+}$ | 3.83 |
|  |  |  |  | $K^{-} \pi^{+} \pi^{-} \pi^{+}$ | 7.49 |
|  |  |  |  | $K^{-} \pi^{+} \pi^{0}$ | 13.9 |
|  |  |  |  | $K_{S} \pi^{+} \pi^{-}$ | 5.4 |
| $D^{*+}$ | 2010.0 | $<5 \times 10^{-21}$ | $<1.5 \times 10^{-6}$ | $D^{0} \pi^{+}$ | 67.7 |
|  |  |  |  | $D^{+} \pi^{0}$ | 30.7 |
| $B^{0}$ | 5279.4 | $1.548 \times 10^{-12}$ | 464 | $D^{*-l^{+} \nu}$ | 4.6 |

Table 1.5: Particle masses and lifetimes especially useful for this analysis [1]. The daughters column shows either the primary decay or a decay which is used in this work.
time. This is similar to watching a film of a particular event backwards. The third operation is Charge Conjugation(C), which simply replaces all particles with their antiparticles (and vice-versa). [10]

The product of all three operations (CPT) is a perfect symmetry of any local field theory by construction: one simply cannot create a renormalizable local field theory without CPT being conserved. The individual symmetries are conserved by both the strong and electromagnetic forces. Initially, physicists had no reason to suspect that any of these symmetries were violated by the weak force, but in 1956, Lee and Yang proposed that the weak force violated parity to explain the $K \rightarrow 2 \pi$ and $3 \pi$, two states with opposite parity. Immediately following this, several new experiments were conducted to confirm this, the first of which was a study of the decay of ${ }^{60} \mathrm{Co}$ with the nuclear spins lined up
parallel in a magnetic field: [8]

$$
\begin{equation*}
{ }^{60} \mathrm{Co} \rightarrow{ }^{60} \mathrm{~N} i^{*}+e^{-}+\bar{\nu}_{e} \tag{1.8}
\end{equation*}
$$

This decay changes the total angular momentum of the nucleus by a single unit of angular momentum, which must be carried away by the electron and antineutrino. The antineutrino is always right-handed which means the electron must be left-handed to conserve angular momentum. Since the charged weak current only couples to the right-handed antineutrino and left-handed electron, there is no "mirror image" reaction. There have been no left-handed antineutrinos directly observed in nature (although recent evidence is pointing to their potential existence [14]), so the parity operator symmetry is maximally broken in this case. The fact that there are left-handed neutrinos (right-handed antineutrinos) but no left-handed antineutrinos (right-handed neutrinos) violates the charge conjugation operator. The combination of parity and charge conjugation (CP) seemed to be a good symmetry since CP takes a left -handed neutrino into a right-handed antineutrino, but even this symmetry was shown to be broken in the kaon system in 1964 by James Cronin and Val Fitch at Brookhaven National Laboratory. CP violation is a very small effect, unlike the maximal symmetry breaking of parity and charge conjugation by the weak force [8].

CP violation has several important implications. The most dramatic is that CP violation represents a difference in the decay properties of matter vs antimatter which offers a possible explanation for the matter/antimatter asymmetry currently observed in the universe. Without an asymmetry in the amount of matter vs antimatter in the
universe, all the matter and antimatter should have ultimately annihilated each other while the universe cooled as it expanded. The resulting universe would be void of all matter and be filled solely with photons [9].

Another implication of CP violation is T violation. Since the combination of CPT must be a perfect symmetry, T must be violated in such a way that that CP violation does not spoil the conservation of CPT. This implies that in certain cases one can find subatomic processes which act differently in the forward time direction than in the backward time direction.

### 1.1.5 Electroweak Interactions

Many of the decays in this analysis are governed by the weak force. The weak force and electromagnetism are united in a single theory in which the electroweak Lagrangian is based on the $\mathrm{SU}(2) \times \mathrm{U}(1)$ symmetry group containing four bosons. Above the spontaneous symmetry-breaking threshold, all four bosons $\left(W^{ \pm}, W^{0}, B^{0}\right)$ are massless and the $\mathrm{SU}(2) \times \mathrm{U}(1)$ symmetry is unbroken. [8]

$$
\begin{equation*}
\mathcal{L}_{\text {electroweak }}=g\left(J^{i}\right)^{\mu} W_{\mu}^{i}+\frac{g^{\prime}}{2}\left(j^{Y}\right)^{\mu} B_{\mu} \tag{1.9}
\end{equation*}
$$

$J^{i}$ is the weak isospin current and $j^{Y}$ is the weak hypercharge current.
Below the symmetry-breaking threshold, the two charged bosons and one linear combination of the neutral bosons acquire mass $\left(W^{ \pm}, Z^{0}\right)$ while the orthogonal combination of neutral bosons remains massless (the photon). The $\mathrm{SU}(2) \times \mathrm{U}(1)$ symmetry is broken, but not completely. There remains a $\mathrm{U}(1)$ symmetry which gives rise to electro-
magnetism and the massless photon. The $W^{ \pm}, Z^{0}$ and photon can be described in terms of the previous fields by:

$$
\begin{gather*}
W_{\mu}^{ \pm}=\sqrt{\frac{1}{2}}\left(W_{\mu}^{1} \mp i W_{\mu}^{2}\right)  \tag{1.10}\\
A_{\mu}=B_{\mu} \cos \theta_{W}+W_{\mu}^{3} \sin \theta_{W}  \tag{1.11}\\
Z_{\mu}=-B_{\mu} \sin \theta_{W}+W_{\mu}^{3} \cos \theta_{W} \tag{1.12}
\end{gather*}
$$

where $\theta_{W}$ is the weak mixing angle, the angle which produces the mass eigenstates of the photon and $Z^{0}$.

The complete Lagrangian can now be written with the more familiar $W^{ \pm}, Z^{0}$ and the photon (with the weak isospin notation, $J_{\mu}=\frac{1}{2}\left(J_{\mu}^{1}+i J_{\mu}^{2}\right)$ ).

$$
\begin{equation*}
\mathcal{L}_{\text {electroweak }}=\frac{g}{\sqrt{2}}\left(J^{\mu} W_{\mu}^{+}+J^{\mu \dagger} W_{\mu}^{-}\right)+e j_{\mu}^{e m} A^{\mu}+\frac{g}{\cos \theta_{W}} J_{\mu}^{N C} Z^{\mu} \tag{1.13}
\end{equation*}
$$

This equation makes use of the following relations:

$$
\begin{gather*}
e=g \sin \theta_{W}=g^{\prime} \cos \theta_{W}  \tag{1.14}\\
e j_{\mu}^{e m}=e\left(J_{\mu}^{3}+\frac{1}{2} j_{\mu}^{Y}\right)  \tag{1.15}\\
J_{\mu}^{N C}=J_{\mu}^{3}-\sin ^{2} \theta_{W} j_{\mu}^{e m} \tag{1.16}
\end{gather*}
$$

Another aspect of the $S U(2) \times U(1)$ symmetry group is the distinction the weak force makes between left and right-handed particles. The charged current (CC) interactions proceed only between left-handed particles, while the neutral current (NC) interactions are predominantly left-handed (right-handed particles participate, but only
through the $j_{\mu}^{e m}$ portion of the NC). The left-handed particles are contained in particle doublets, while right-handed particles form singlet states.

The weak eigenstates of the quarks differs from the mass eigenstates. It is necessary to describe the down type weak eigenstate quarks as a superposition of down type mass eigenstate quarks. One could have used the up type quarks instead of down type, but there would be no difference since the absolute phases of quark wave functions are not observable. More complicated mixing schemes can be employed but can always be reduced down to this choice of phases. The quark mixing is given by the fundamental Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix: [8]

$$
\left(\begin{array}{c}
d^{\prime}  \tag{1.17}\\
s^{\prime} \\
b^{\prime}
\end{array}\right)=\left(\begin{array}{ccc}
V_{u d} & V_{u s} & V_{u b} \\
V_{c d} & V_{c s} & V_{c b} \\
V_{t d} & V_{t s} & V_{t b}
\end{array}\right)\left(\begin{array}{c}
d \\
s \\
b
\end{array}\right)
$$

As we will see, form factor extraction is critical for determination of two of the entries of this basic matrix.

### 1.2 The Role of Form Factors in Semileptonic Decays

One of the most useful and clean decay mode categories of the $B$ meson that particle physicists have access to is that of semileptonic decays. And though all the above formalism in Section 1.1.3 is general and applicable to any type of decay process, in fact in semileptonic decays we are able to evade many of the hadronic uncertainties that plague
processes in which there are no leptons in the final state. This is because in fully hadronic processes, nonperturbative QCD amplitudes play an even larger role in the decay, making them less amenable to a perturbative expansion analysis.

But semileptonic decays require at least one of the decay vertices to be weak, immediately making at least that part of the decay amplitude open to a perturbative expansion. In these types of decays, the virtual $W$ boson which changes the flavor of the b-quark into an up or charm quark creates a lepton-neutrino pair instead of a quark antiquark doublet as in hadronic decays. Thus, we may exploit the relative lack of hadronic uncertainties in this mode of decay (compared to decays to fully hadronic final states) to give us clearer and less ambiguous information about many aspects of both the weak and strong interactions.

The functions that parameterize our ignorance about the calculationally difficult non-perturbative QCD processes between the quarks in the initial and final states are called "form factors", and quantitative determination of these can be useful for several physics motivations as we will see in the following sections. What is important to note is that because the virtual $W$ in semileptonic processes decays only to leptons, these can have no (tree-level) interactions with the other quarks in the decay so that the process can be broken down into two pieces and the form factors actually determined numerically and then later applied in other fully hadronic processes.

As examples of things we can obtain information about form determination of the form factors, we will see that we may learn about elements in the fundamental

CKM matrix (which consists of the coupling strengths between up-type and down-type quarks as described in Section 1.1.5), about the internal wavefunction structure of the $B$ mesons themselves, and about allowed Quantum Chromodynamics (QCD) inspired Heavy Quark Effective Theory (HQET) models for these decay processes.

### 1.2.1 Heavy Quark Effective Theory (HQET)

One benefit of studying semileptonic decays in modes where the $b$ quark transmutes into the $c$ quark is that we may study HQET directly. HQET[16] is the framework in which we exploit heavy quark symmetry (HQS), which in simple terms is just that symmetry whereby we can think of the system outside of the central heavy quark which is a complex non-perturbative quantity consisting of the light valence quark, and also all exchanged sea quarks and gluons, and is thus normally referred to collectively as the "light degrees of freedom" (or more colloquially, the "brown muck") - as being "unconcerned" with the change of flavor of the central heavy quark as it transforms from a $b$ to a $c$ quark. As long as the momentum of this quark relative to the accompanying cloud of "brown muck" doesn't vary appreciably, the muck continues to exchange gluons with the heavy quark informing both the muck and the heavy quark that they are still bound together as one system and none of the interactions have changed - i.e., under pure HQS, QCD is absolutely flavor-blind in this transformation.

### 1.3 Determination of CKM element $V_{c b}$

Semileptonic decay measurements in transitions of $b$ to $c$ quarks always yield a rate which is proportional to a product of $\left|V_{c b} \times \mathcal{F}_{D^{*}}(1)\right|^{2}$. Here $V_{c b}$ is one of the CKM matrix elements as it is shown in Section 1.1.5, specifically the row 2, column 3 entry giving the coupling between the $c$ and $b$ type quarks. The second factor is a generalized form factor, which encodes our ignorance about the detailed QCD dynamics between the $b$ and $c$ quarks in the decay process (its argument is $w$, a kinematic property of the event, described in detail in Section 3.1).

One theoretical framework for parameterizing the form factor is HQET. In this framework $\mathcal{F}_{D^{*}}(1)$ can be parameterized into a form normally referred to as the IsgurWise (IW) function [12] evaluated for this decay at zero recoil of the $D^{*}$ meson (technically, $\mathcal{F}_{D^{*}}(1)$ is composed of the three separate form factors involved in this decay, but it is often interchangeably referred to in the singular as "the" form factor for the decay process). Thus, once the rate is measured, this form factor must be determined from elsewhere in order to extract a value for $V_{c b}$ using this exclusive decay reconstruction technique - i.e. schematically, $V_{c b}=R A T E / \mathcal{F}_{D^{*}}(1)$. The more accurately we can determine this form factor function, the more we can reduce the errors in the $V_{c b}$ extraction.

The value of $\mathcal{F}_{D^{*}}(1)$ in the infinitely heavy quark mass limit is unity, but there are finite-mass and higher order $\alpha_{s}$ corrections that modify it to $\sim 0.93$ [16]. The behavior of $\mathcal{F}_{D^{*}}(w)$ as a function of $w$ is dependent on the form factors that this total function is composed of and is what we are determining in this measurement, and since the am-
plitude goes to zero at $w=1$, some extrapolation of $\mathcal{F}_{D^{*}}(w)$ to the $w=1$ point is needed which requires an understanding of the slope as this point is approached (in the color version of this document, the color purple will be used to identify any of the four kinematic variables, $w$ being one, along with the three angular ones introduced in the next chapter; see Appendix A.1).

The current numerical values quoted by BaBar using the CLEO-determined form factors [4] for $\mathcal{F}(1)\left|V_{c b}\right|$ are [39]: $10^{-4} \times\left[35.52 \pm 0.25(\right.$ stat $) \pm 1.25($ syst $\left.) \pm 0.85\left(F F^{\prime} s\right)\right]$ (and $\mathcal{F}=0.93 \pm 0.03$ from HQET calculation [20], where the error is from theoretical extrapolation), so it can be seen that the form factor contribution to the total error is large.

### 1.4 Determination of CKM element $V_{u b}$

$V_{u b}$ extraction relies on reconstruction of $b \rightarrow u$ quark states, which are suppressed a priori from $b \rightarrow c$ quark states by a factor of $\left|\frac{V_{u b}}{V_{c b}}\right|^{2} \sim 0.01$, and thus it is clear that all $b \rightarrow c$ decays will cause enormous backgrounds to $V_{u b}$ extraction, and the properties. One place this is manifested is in the $p_{l}^{*}$ spectrum which depends strongly on the form factor inputs to the MC model. Only after the $p_{l}^{*}$ spectrum is correctly determined and subtracted can the extraction of $V_{u b}$ proceed without worry that the ultimate measurement is biased [38].

### 1.5 This Measurement

We present in this dissertation the results of a measurement of the three semileptonic form factors involved in the decay $B^{0} \rightarrow D^{*} \ell \nu$ where we mean here either of the two charge conjugate modes $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ or $\bar{B}^{0} \rightarrow D^{*+} \ell^{-} \bar{\nu}$, but will generally write simply $B^{0} \rightarrow D^{*} \ell \nu$ and understand it to mean either. Similarly, unless we specify otherwise, when we write any given meson, we will mean either charge conjugate state (i.e., by " $B^{0}$ " we will mean either neutral $B$ meson, i.e., either the $B^{0}$ or the $\bar{B}^{0}$, by " $D^{0 \text { " }}$ we will mean either the $D^{0}$ or the $\bar{D}^{0}, " D^{\prime \prime}$ will denote either a $D^{+}$or a $D^{-}$), and finally also we mean by " $\ell$ " here either of the two light charged leptons (i.e., electron or muon).

The $D^{*}$ is reconstructed for this work in the mode $D^{*} \rightarrow D^{0} \pi^{ \pm}$, and the $D^{0}$ in the mode $D^{0} \rightarrow K^{-} \pi^{+}$.

This measurement uses the Babar Run 1+Run 2 Years 2000-2002 data set, which yield altogether approximately $85 \times 10^{6} B \bar{B}$-pairs.

The method of the measurement is to perform an unbinned maximum likelihood fit in the four kinematic variables that describe the decay for the three parameters $R_{1}, R_{2}$, and $\rho^{2}$.

The previous results for these quantities were obtained from the CLEO detector at Cornell, and were: $R_{1}=1.18 \pm 0.30 \pm 0.12$ and $R_{2}=0.71 \pm 0.22 \pm 0.07$ and form factor slope was found to be $\rho^{2}=0.91 \pm 0.16 \pm 0.06[4]$.

## Chapter 2

## Analysis Method Overview

Before we delve into the formalism involved in this process, we will give an outline of how the analysis is carried out. This will help to motivate and map out the next several chapters of this rather involved analysis chain.

In an ideal world, one would simply take an unbiased sample of $D^{*} \ell \nu$ events, record the directions and energies of all the visible particles, extract the kinematic variables (see Sections 3.1 and 3.2) from each of these events, and fit the resulting multivariate four-dimensional data distribution to the theoretical expectation (given in Section 3.3.2) to determine the three form factors and compare them to theoretical predictions (given in Section 3.6). Unfortunately, the real world has acceptance effects, smearing, backgrounds, and no ability to detect the neutrino, among other complications. So this basic recipe serves as the framework for this measurement, but each step requires a considerable amount of work to actually accomplish:

Step 1. Extract sample of $D^{*} \ell \nu$ events: The first step is to create a sample of candidate $D^{*} \ell \nu$ events in as pure a form as possible. This is done by the standard method of making candidate events by combining basic tracks and energy clusters, at each stage of this process making a variety of cuts to keep as efficient and pure a sample as possible (these cuts and their effects are covered in detail in Sections 5.2.7-5.2.11). These cuts suppress the primary backgrounds which are: inclusive $B \rightarrow D^{*} \ell \nu X$ decays where $X$ might be e.g. an extra pion that is missed in the reconstruction process, and backgrounds due to misreconstruction of tracks where a given event is only mimicking an actual $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ event (e.g. misreconstruction of any of the daughter particles in the $D^{*} \rightarrow D^{0} \pi, D^{0} \rightarrow K^{-} \pi^{+}$chains), and continuum events $(c \bar{c})$. These various different types of backgrounds require different types of cuts to suppress them during the reconstruction process which are detailed as we describe the reconstruction itself.

The several stages of the candidate reconstruction are covered as so:

- First a candidate $D^{0}$ is created from all potential kaon and pion candidates in the $D^{0} \rightarrow K^{-} \pi^{+}$mode (see Section 5.2.2).
- Next the four-vector of another candidate pion is added to that of the $D^{0}$ to create a candidate $D^{*}$ particle (as covered in Section 5.2.6).
- Last a candidate lepton (electron or muon) is added to the candidate $D^{*}$ to form a final candidate $D^{*} \ell$ object (as also covered in Section 5.2.6).

Beyond this, there are final cuts made on the $D^{*} \ell \nu$ sample itself to winnow out as many of the remaining backgrounds as possible, keeping the best attainable signal to noise ratio. We then have our final sample of all candidates, and any remaining background events are removed in the background subtraction process which is part of the fit as covered in the next step.

## Step 2. Fit the sample of $D^{*} \ell \nu$ candidate events to the theoretical PDF:

- We do a fit by first accounting for smeared directions and momenta of all the detectable particles into the fitting PDF. We lump together here parent $B^{0}$ meson directional uncertainty due to the missing neutrino, acceptance, and detector and kinematic variable resolution effects. Subtraction of the background is also part of this fit process. Dealing with all these issues and correcting for them is done in several substeps (this is dealt with primarily in Chapter 6):
- First we calculate the event kinematic variables in a frame which is chosen to provide the best resolution possible relative to what they would actually be in the exact $B$ rest frame (covered in Section 6.3.2).
- Then we use the unique moments expansion method (as described in Sections 6.1-6.2) to correct for acceptance effects. This method effectively builds a model of the acceptance from a full MC simulation and then unfolds the data using this model.
- We also subtract background as a part of the fitting process. Mathematically, this is dividing a product of the PDF evaluated at all points by a
product of the PDF evaluated at each background event, so that this distribution of background events is removed entirely (up to statistical effects and MC / data differences from the quantity that is ultimately being fitted). An exact description of how this is done is given in Section 6.4:
- The fit yields the two form factor ratios $R_{1}$ and $R_{2}$, and the form factor slope $\rho^{2}$ (as described in Section 3.4).
- After the fit, we must ex post facto adjust each form factor for the part of the resolution which was not taken into account into the fitting PDF (Section 6.6).
- Goodness-of-Fit: After the final form factor values are extracted, we estimate some measure of goodness-of-fit. Though multi-dimensional goodness-of-fit for strongly correlated parameters is notoriously difficult to estimate for unbinned maximum likelihood fits, we found through our validation tests that a quite intuitively simple $\frac{\chi^{2}}{\text { dof }}$-based method (as described in Section 6.8) gives us a trustworthy estimate of the goodness-of-fit.


## Step 3. Determine All Errors:

The determination of all the errors and biases associated with the kinematic variable measurements and the subtraction of the various backgrounds, as well as with those associated with limited MC statistics etc. is one of the most involved aspects of the form factor measurement, and the final step in obtaining the experimental result. This is primarily covered by Chapter 8 in several steps:

- Error can be broken down into four separate major subtypes:
- Standard data statistical: This is the error Minuit provides us when the fit is run.
- Extra statistical error: The error due to the way in which we handle the imperfect resolution becomes an extra data statistical error (the discussion of why this occurs is initially begun in Section 6.6 where it is relevant, and the formulae for its evaluation are given later in Section 8.1).
- MC statistical error: The MC statistical error estimation is rather unusual and specific to our methods of acceptance correction and background subtraction (this is covered in Sections 8.2 and 8.3).
- Standard Systematic: Other systematic errors are evaluated in standard ways (this is covered in the rest of Chapter 8).

The final result is a measurement of the three form factors, plus all associated errors, statistical and systematic. Along with these is an estimated goodness-of-fit measure.

Conclusions and prospects: Finally, we give our final results in Chapter 9 and also projections for what results could be achieved with the full BaBar dataset as envisioned in the foreseeable future.

## Chapter 3

## Formalism

### 3.1 Quark-level Description

We now describe the relevant quantities needed for this form factor analysis, beginning with the basic quark-level diagram for this process, shown in Figure 3.1.


Figure 3.1: Quark level diagram of $B \rightarrow D^{*} \ell \nu$ decay.

Here the $\bar{d}$ is a spectator quark, and the $b$ quark decays weakly through the $W-b-c$ vertex with the associated CKM element $V_{c b}$ into a $c$ quark and a virtual $W$.

This $W$ then further decays to a lepton-antineutrino pair. The $c$ and $\bar{d}$ hadronize into an outgoing $D^{*}$ so that we finally have the three-body decay $B^{0} \rightarrow D^{*} \ell \nu$. The quantity $q^{2} \equiv\left(p_{B}-p_{D^{*}}\right)^{2}$ (where $p_{B}$ and $p_{D^{*}}$ are the 4-momenta of the $B$ and $D^{*}$, respectively) is called the "momentum transfer", and varies from $0 \rightarrow \sim 10.7(\mathrm{GeV} / c)^{2}$ in this decay.

The momentum transfer $q^{2}$ is linearly related to another Lorentz invariant which is useful to characterize the decay:

$$
\begin{equation*}
w \equiv \frac{p_{B} \cdot p_{D^{*}}}{M_{B} M_{D^{*}}}=v_{B} \cdot v_{D^{*}}=\frac{M_{B}^{2}+M_{D^{*}}^{2}-q^{2}}{2 M_{B} M_{D^{*}}} \tag{3.1}
\end{equation*}
$$

where $p$ is the 4 -momentum and $v$ is the 4 -velocity (i.e., $\frac{p}{m}$ ) for the respective particle. In the rest frame of the decaying $B^{0}, w$ reduces to:

$$
\begin{equation*}
w \rightarrow \frac{\left(M_{B}, 0\right) \cdot\left(E_{D^{*}}, \wp_{D^{*}}\right)}{\left(M_{B} M_{D^{*}}\right)}=\frac{E_{D^{*}}}{M_{D^{*}}} \tag{3.2}
\end{equation*}
$$

where $\wp_{D^{*}}$ is the magnitude of the three-momentum in this frame.
To get a more physical feeling for these variables, we may note that using $p_{B}=p_{D^{*}}+p_{l}+p_{\nu}$, we find $q^{2}$ is equivalently $\left(p_{l}+p_{\nu}\right)^{2}$, the invariant mass of the virtual $W$. Higher $q^{2}($ lower $w)$ thus corresponds to a higher mass of the virtual $W$, which at the two-body decay level, implies a lower "kick" to the $D^{*}$ - in fact, $w$ is just the relativistic boost factor $\gamma$ of the $D^{*}$ in the $B$ rest frame, as can be seen from eq.8.21.

Another function that varies with $w$ is the quantity $\xi(w)$, called the Isgur-Wise (I-W) form factor[15], which can be understood as the wavefunction overlap of the initial $B^{0}$ and the final state $D^{*}$ during this transition process. This can be seen to fall with

| Configuration | $q^{2}$ | $w$ | $\xi(w)$ |
| :---: | :---: | :---: | :---: |
| Stationary $D^{*}$ | $\sim 10.7(\mathrm{GeV} / c)^{2}$ | 1.000 | 1 (maximal) |
| Max. boosted $D^{*}$ | $m_{l}^{2} \sim 0$ | 1.50377 | $\sim .8$ (minimal) |

Table 3.1: Ranges of $w, q^{2}$, and $\xi(w)$.
increasing $w$ due to the necessity of exchange of more soft gluons when the daughter quark has more velocity relative to the parent quark (i.e., in this case the gluon exchanges must communicate to the daughter quark that it needs to "move" relative to its mother, where no exchange is necessary when it continues with the same velocity). Thus the I-W function is maximal at zero recoil. Table 3.1 shows the extremal configurations for this decay, and the corresponding values for $w$ and the I-W form factor.

### 3.2 Angular Kinematic Variables

The quark level diagram becomes, upon hadronization of the quarks, the decay of the $B^{0}$ to $D^{*} \ell \nu$, and we choose to reconstruct the mode where the $D^{*}$ decays further to a $D^{0} \pi^{ \pm}$(this occurs about $2 / 3$ of the time, the other $1 / 3$ of the time the $D^{*}$ goes to a $D^{ \pm} \pi^{0}$, with a tiny fraction (about $2 \%$ ) going to $D^{ \pm} \gamma$ ). We choose to reconstruct the $D^{0} \pi^{ \pm}$mode because detecting charged pions with good track direction and momentum resolution is much easier to do for charged tracks than for neutrals.

The outgoing particles which are seen in the detector are shown in Figure 3.2 (though of course the neutrino is lost), and the $D^{0}$ is reconstructed through its observed
daughters, as we will see.


Figure 3.2: Meson level diagram of $B \rightarrow D^{*} \ell \nu$ decay (for the decay we are analyzing, the $B$ is neutral, the $D^{*}$ charged, as is the $\pi$, and the $D$ daughter meson is neutral).

This figure defines three angles, which, along with $w$ form the four independent kinematic variables we use to fully characterize this decay in the rest frame of the $B^{0}$. Two of these angles are taken in their cosine form and one directly as kinematic variables:

- $\cos \theta_{\ell}$ : the cosine of the included angle between the direction of the lepton boosted into the virtual $W$ rest frame, and the direction of the virtual $W$ in the $B^{0}$ rest frame; ranges from $-1 \rightarrow 1$.
- $\cos \theta_{V}$ : the cosine of the included angle between the the direction of the $D^{0}$ boosted into the $D^{*}$ rest frame, and the direction of the $D^{*}$ in the B rest frame; ranges from $-1 \rightarrow 1$
- $\chi$ : the azimuthal angle between the half plane formed by the $D^{*}-D^{0}$ system and the half plane formed by the $W-\ell$ system; ranges from $0 \rightarrow \pi$.


### 3.3 Four Dimensional Probability Distribution Function

In order to derive the differential decay rate (which is a form of a "probability distribution function", or PDF, and is commonly experimentally referred to this way) as a function of $w$ we start with the usual matrix element for a semileptonic decay $M_{Q \bar{q}} \rightarrow X_{q^{\prime} \bar{q}} \ell^{-} \bar{\nu}$, which can be written as the product of a hadronic and a leptonic current:

$$
\begin{align*}
\mathcal{M}\left(M_{Q \bar{q}} \rightarrow X_{q^{\prime}} \ell^{-} \bar{\nu}\right) & =\left\langle X_{q^{\prime} \bar{q}}\right| \bar{q}^{\prime} \frac{-i g}{2 \sqrt{2}} V_{q^{\prime} Q} \gamma_{\mu}\left(1-\gamma_{5}\right) Q\left|M_{Q \bar{q}}\right\rangle \\
& \times P^{\mu \nu}(q) \bar{u}_{\ell} \frac{-i g}{2 \sqrt{2}} \gamma_{\nu}\left(1-\gamma_{5}\right) v_{\nu} \tag{3.3}
\end{align*}
$$

where $Q$ is the annihilation operator for a quark Q . The $W$ propagator is given by

$$
\begin{equation*}
P^{\mu \nu}(q)=\frac{i\left(-g^{\mu \nu}+q^{\mu} q^{\nu} / M_{W}^{2}\right)}{q^{2}-M_{W}^{2}} \tag{3.4}
\end{equation*}
$$

If the energy of the virtual $W, q^{2}$, is much less than $M_{W}$ it is convenient to write $P^{\mu \nu}(q) \approx i \frac{g^{\mu \nu}}{M_{W}^{2}}$.

We now rewrite the matrix element for this process as:

$$
-i \mathcal{M}=-i \frac{G_{F}}{\sqrt{2}} V_{c b} L_{\mu} H^{\mu}
$$

We write the simple lepton tensor as $L_{\mu}=\bar{l} \gamma_{\mu}\left(1-\gamma_{5}\right) \nu$, while the hadronic tensor must be parameterized completely generally in terms of three helicity amplitudes (for leptons with negligible mass, for massive leptons there are in general four nonzero helicity amplitudes that contribute). [16] These helicity amplitudes are pure functions of
$w$ because they are defined to be quark-level functions which encode our ignorance about what is happening non-perturbatively with QCD at the quark level, and thus have no knowledge of the angles involved in the hadronization process that happens afterwards.

Taking into account Lorentz invariance, and using the available four vectors (the $D^{*}$ polarization $\epsilon^{\alpha}$, the $B^{0} 4$-momentum $p^{\mu}$, and the $D^{*} 4$-momentum $k^{\nu}$ ), the hadronic tensor can be parameterized completely generally in terms of four form factors (in the color version of this document, the color red will be used to identify form factors when they appear, including all HQET forms $H$ introduced below, as well as the older forms $A, V$ used in this form, see Appendix A.1):

$$
\begin{aligned}
H^{\mu} & =\left\langle D^{*+}(k, \epsilon)\right| \bar{c} \gamma^{\mu}\left(1-\gamma_{5}\right) b\left|\bar{B}^{0}(p)\right\rangle= \\
& \frac{2 i \epsilon^{\mu \nu \alpha \beta}}{M_{B}+M_{D^{*}}} \epsilon_{\nu}^{*} k_{\alpha} p_{\beta} V\left(q^{2}\right)-\left(M_{B}+M_{D^{*}}\right) \epsilon^{* \mu} A_{1}\left(q^{2}\right)+ \\
& \frac{\epsilon^{*} \cdot p}{M_{B}+M_{D^{*}}}(p+k)^{\mu} A_{2}\left(q^{2}\right)-\frac{\epsilon^{*} \cdot p}{M_{B}+M_{D^{*}}}(p-k)^{\mu} A_{3}\left(q^{2}\right) .
\end{aligned}
$$

(the letters $V$ and $A$ stand for the underlying vector and axial vector couplings). $A_{3}$ can be neglected in the limit of small lepton mass relative to the meson masses in the process, as can be easily seen by just contracting the lepton and hadron tensors, in which case it gets a coefficient proportional to $\frac{m_{l}^{2}}{q^{2}}$ (so this case holds for both the e and $\mu$, but not the $\tau$ lepton), thus we drop it henceforth (except below where we discuss explicitly the $m_{\ell}^{2}$ dependence of the PDF).

### 3.3.1 HQET helicity amplitudes

The Lorentz structure of the $D^{*} l \nu$ decay amplitude can be expressed in terms of three more physically intuitive amplitudes ( $H_{+}, H_{-}$and $H_{0}$ ), which correspond to the three allowed polarization states of the $D^{*}$ (two transverse and one longitudinal). These amplitudes are related to the axial and vector form factors as follows:

$$
\begin{array}{r}
H_{+}\left(q^{2}\right) \equiv-\left(M_{B}+M_{D^{*}}\right) A_{1}\left(q^{2}\right)+2 \frac{\wp_{D^{*}} M_{B}}{M_{B}+M_{D^{*}}} V\left(q^{2}\right), \\
H_{-}\left(q^{2}\right) \equiv-\left(M_{B}+M_{D^{*}}\right) A_{1}\left(q^{2}\right)-2 \frac{\wp_{D^{*}} M_{B}}{M_{B}+M_{D^{*}}} V\left(q^{2}\right), \\
H_{0}\left(q^{2}\right) \equiv-\frac{1}{2 M_{D^{*} \sqrt{q^{2}}}}\left(A_{1}\left(q^{2}\right)\left(M_{B}+M_{D^{*}}\right)\left(M_{B}^{2}-M_{D^{*}}^{2}-q^{2}\right)-\right. \\
\left.4 \frac{M_{B}^{2} \wp_{D^{*}}^{2}}{M_{B}+M_{D^{*}}} A_{2}\left(q^{2}\right)\right)
\end{array}
$$

To obtain the full differential cross section decay rate, we contract the lepton and hadron tensors, and carry out the phase space integrations, to find the PDF with respect to the four relevant kinematic variables.

### 3.3.2 Simplified PDF

From Appendix B, we are justified for the moment in assuming no CP violation and fully neglecting the terms in the PDF that depend on $m_{\ell}$, thus taking the three helicity amplitudes to be real, and setting $m_{\ell} \rightarrow 0$, we find the simplified form of the PDF that we
will be working with henceforth:

$$
\begin{align*}
\frac{d \Gamma\left(B \rightarrow D^{*} \ell \nu\right)}{d q^{2} d \cos \theta_{\ell} d \cos \theta_{V} d \chi}= & \frac{3 G_{F}^{2}\left|V_{c b}\right|^{2} \wp_{D^{*} q^{2}}}{8(4 \pi)^{4} M_{B}^{2}} \\
& \left\{H_{+}{ }^{2}\left(1-\cos \theta_{\ell}\right)^{2} \sin ^{2} \theta_{V}+H_{-}^{2}\left(1+\cos \theta_{\ell}\right)^{2} \sin ^{2} \theta_{V}\right. \\
& +4 H_{0}{ }^{2} \sin ^{2} \theta_{\ell} \cos ^{2} \theta_{V}-2 H_{+} H_{-} \sin ^{2} \theta_{\ell} \sin ^{2} \theta_{V} \cos (2 \chi) \\
& -4 H_{+} H_{0} \sin \theta_{\ell}\left(1-\cos \theta_{\ell}\right) \sin \theta_{V} \cos \theta_{V} \cos \chi  \tag{3.5}\\
& \left.+4 H_{-} H_{0} \sin \theta_{\ell}\left(1+\cos \theta_{\ell}\right) \sin \theta_{V} \cos \theta_{V} \cos \chi\right\}
\end{align*}
$$

Where $\wp_{D^{*}}$ is the magnitude of the $D^{*}$ momentum.

### 3.4 Helicity Amplitudes and PDF Factors

[ Note: in this section, two other parameterizations will be used for the form factors in addition to the modern $H_{+}, H_{-}, H_{0}$ helicity form: the older $V, A_{1}, A_{2}$ form, and the phase-space reduced $\tilde{H}_{+}, \tilde{H}_{-}, \tilde{H}_{0}$ form. The important thing to remember throughout the transformations between these forms is that the three experimental parameters we will be measuring in this analysis are the two helicity-amplitude ratios $R_{1}, R_{2}$, and the slope $\rho^{2}$, which collectively are termed the three measured "form factors" for this analysis. ]

Because certain theoretical uncertainties drop out in the ratio, and the ratios are easier to predict than the absolute form factors, we now define two further quantities:

$$
\begin{aligned}
R_{1}\left(q^{2}\right) & \equiv\left[1-\frac{q^{2}}{\left(M_{B}+M_{D^{*}}\right)^{2}}\right] \frac{V\left(q^{2}\right)}{A_{1}\left(q^{2}\right)}, \\
R_{2}\left(q^{2}\right) & \equiv\left[1-\frac{q^{2}}{\left(M_{B}+M_{D^{*}}\right)^{2}}\right] \frac{A_{2}\left(q^{2}\right)}{A_{1}\left(q^{2}\right)} .
\end{aligned}
$$

(in the color version of this document, the color green will be used to identify these ratios of form factors when they appear, see Appendix A.1).

We now introduce another parameterization of the form factors motivated by HQET models:

$$
\begin{equation*}
h_{A_{1}}(w)=R^{*}\left[1-\frac{q^{2}}{\left(M_{B}+M_{D^{*}}\right)^{2}}\right]^{-1} A_{1}\left(q^{2}\right) \tag{3.6}
\end{equation*}
$$

with:

$$
\begin{equation*}
R^{*} \equiv \frac{2 \sqrt{M_{B} M_{D^{*}}}}{\left(M_{B}+M_{D^{*}}\right)} \tag{3.7}
\end{equation*}
$$

Because recent predictions are made based on HQET models, we may further write in this parameterization the polarization amplitudes as explicit functions of the HQET parameters defined above:

$$
\left.\begin{array}{l}
\tilde{H}_{+}=\left[\frac{\sqrt{\left(1-2 w r+r^{2}\right)}\left(1-\sqrt{\frac{w-1}{w+1}} R_{1}\right)}{(1-r)}\right] \\
\tilde{H}_{0} \\
\tilde{H}_{-}=\left[1+\frac{\sqrt{(w-1)\left(1-R_{2}\right)}}{1-r}\right]  \tag{3.9}\\
(1-r)
\end{array}\right]
$$

With the translation between the $H$ forms in eq. 3.5 and the $\tilde{H}$ forms of eq. 3.8 being:

$$
\begin{equation*}
H_{i}=\frac{h_{A_{1}}(w)(w+1) \tilde{H}_{i}}{\sqrt{q^{2}}}, \wp_{D^{*}}=M_{D^{*}} \sqrt{w^{2}-1}, r \equiv \frac{M_{D^{*}}}{M_{B}} \tag{3.10}
\end{equation*}
$$

Going back to $h_{A_{1}}$ from eq. 3.6, using the connection between $A_{1}$ and $\rho_{A_{1}}^{2}$ : we may expand this form factor in a Taylor series around the point $w=1$ as so:

$$
\begin{equation*}
h_{A_{1}}(w)=h_{A_{1}}(1)\left[1-\rho_{A_{1}}^{2}(w-1)+\ldots\right], \quad h_{A_{1}}(1.0)=1.0 \tag{3.11}
\end{equation*}
$$

where $\rho_{A_{1}}^{2}$ is called the 'slope' of the form factor; for small higher order coefficients ("curvature terms") this is a good approximation over the entire allowed $w$ range $1 \rightarrow \sim 1.504$. HQET corrections (of order $\alpha_{s}$ and $\frac{1}{m_{Q}}$ ) modify $h_{A_{1}}$ from the HQET limit of unity (the color green will be used to identify also this slope $\rho^{2}$ when it appears, as it is our third form factor).

We have now formally introduced the triplet of parameters that we will be measuring in this analysis: $R_{1}, R_{2}$ and $\rho^{2}$ (for brevity, we will henceforth drop the $A_{1}$ subscript on the slope $\rho^{2}$, as we have here).

In Appendix C we show results of fitting generator-level MC with this theoretical PDF, as a simple initial verification of our MC-generation and fitting tools.

### 3.5 Contributions of $H_{[+, 0,-]}^{2}$

Because the $c$ quark inherits the helicity of the $b$ quark upon decay, the V-A interaction implies that the $H_{-}(w)^{2}$ term dominates except near the high endpoint of $w$ where only $H_{0}(w)^{2}$ can contribute.

We can see this in Figure 3.3 (vertical scale arbitrary) which shows the relative contribution of the three helicity states of the $D^{*}$ as a function of $w$ (this is shown for the CLEO values of the form factors [4], though the curves don't change dramatically for nearby other values of the form factors).

These arise as a consequence of the V - A interaction: taking the decay $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ the $b$ quark decays to a left-handed $c$ quark in the limit of massless quarks.

So though a nearly stationary decay $D^{*}$ is unpolarized (this is the configuration shown in Figure 3.4), and thus is a combination of all three states equally, as the velocity of the ejected $c$ quark becomes higher (or equivalently, as $w$ increases), and the $c$ quark approaches more closely a massless state, it becomes preferentially left-handed, and combining its helicity of $-1 / 2$ with the light spectator quark (which has an equal probability
of being helicity $1 / 2$ or $-1 / 2$ ) leads to a helicity of 0 or -1 for the resultant $D^{*}$.


Figure 3.3: Various helicity component contributions to the PDF (plots made with the CLEO values of the form factors).

But we also have Figure 3.5 which shows that as $w$ goes maximal, the negative lepton and antineutrino combine into a helicity zero state, forcing also the $D^{*}$ into the helicity zero projection, leading to suppression of the helicity -1 amplitude.


Figure 3.4: Minimal $(w=1.000)$ configuration
(The configuration of Figure 3.5 is somewhat analogous to the two-body process of a pseudoscalar $\pi^{-}$decaying almost purely to a negatively charged lepton and


Figure 3.5: Maximal ( $w=1.504$ ) configuration.
antineutrino, the preferential states of which are LH and RH respectively in the limit of zero lepton mass. This is what forces the $\pi$ to decay to a $\mu$ vs. an electron because the original pion has spin zero and the electron has very little RH component in it, relative to the $\mu$. In this case though, instead of the decay being suppressed, the $D^{*}$ is forced into the helicity zero state.)

### 3.6 Form Factor Predictions

For infinitely massive heavy $b$ and $c$ quarks, we expect $R_{1}=R_{2}=1.0$, but these are modified by both perturbative $\alpha_{s}$ and non-perturbative $\left(\frac{\Lambda_{\mathrm{QCD}}}{m_{x}}\right)$ corrections.

Calculating higher order loop corrections to the form factors yields an expansion of the form:

$$
\begin{aligned}
& R_{1}(w)=1+\left[\alpha_{s}(\ldots)+\alpha_{s}^{2} \beta_{0}(\ldots)\right]+\left(\frac{\Lambda_{\mathrm{QCD}}}{m_{x}}\right)(\ldots), \\
& R_{2}(w)=1+\left[\alpha_{s}(\ldots)+\alpha_{s}^{2} \beta_{0}(\ldots)\right]+\left(\frac{\Lambda_{\mathrm{QCD}}}{m_{x}}\right)(\ldots)
\end{aligned}
$$

The ellipses after the $\alpha_{s}$ terms have been calculated perturbatively up to an order which gives confidence that they are accurate to just a few percent (see [18] for this and the following paragraphs of this section).

The coefficients of the $\left(\frac{\Lambda_{\mathrm{QCD}}}{m_{x}}\right)$ terms are combinations of quantities called "subleading Isgur-Wise functions" (SIWF's), of which there are four for the decay $B^{0} \rightarrow D^{*} \ell \nu$ (usually denoted $\chi_{1,2,3}$ and $\eta$ ).

These quantities are much less certain than the $\alpha_{s}$ corrections, as they must be calculated using non-perturbative techniques (e.g. from models, sum rules, or the lattice), and the last CLEO measurement included the full range allowed by various models. It did not restrict the ranges allowed for the SIWF's significantly.

However the current measurement narrows the allowed range down much more precisely as we shall see. And since the entire method of extracting $V_{c b}$ using an HQET
expansion, as well as other predictions, relies on believing the corrections to HQET are predictable, we are testing HQET validity at a basic level with this measurement.

Different models in the HQET framework evaluate the SIWF correction terms differently, ending in varying predictions for $R_{1}$ and $R_{2}$. As examples, Neubert [16] based on early work with collaborators in the early 90's predicts:

$$
\begin{align*}
& R_{1}(w)=1.35-0.22(w-1)+0.09(w-1)^{2}  \tag{3.12}\\
& R_{2}(w)=0.79+0.15(w-1)-0.04(w-1)^{2} \tag{3.13}
\end{align*}
$$

More recently, Caprini et.al. [19] using spectral functions, dispersion relations, and heavy quark symmetry to evaluate the non-perturbative terms predict:

$$
\begin{align*}
& R_{1}(w)=1.27-0.12(w-1)+0.05(w-1)^{2}  \tag{3.14}\\
& R_{2}(w)=0.80+0.11(w-1)-0.06(w-1)^{2} \tag{3.15}
\end{align*}
$$

Ligeti and Grinstein [18] using similar HQET machinery predict:

$$
\begin{align*}
& R_{1}(w)=1.25-0.10(w-1)  \tag{3.16}\\
& R_{2}(w)=0.81+0.09(w-1) \tag{3.17}
\end{align*}
$$

Whereas all the above HQET-based predictions predict form factor values within a fairly narrow range, older predictions relying on guessed potential models vary widely, e.g. Close \& Wambach using a simple quark model predict [17]:

$$
\begin{align*}
& R_{1}(w)=1.15-0.07(w-1),  \tag{3.18}\\
& R_{2}(w)=0.91+0.04(w-1) \tag{3.19}
\end{align*}
$$

In this work we assume that the coefficients of the $(w-1),(w-1)^{2}$ are small enough that $R_{1}$ and $R_{2}$ are constant over the entire range of $w$ - in effect, we determine a $w$-averaged value of them ${ }^{1}$.

The slope $\rho^{2}$ from eq.(3.11) does not enter the above predictions but must be separately determined from sum rules or lattice calculations[20].

[^2]
## Chapter 4

## $B_{A} B_{A R}$ Detector Overview

### 4.1 Detector Overview

Though the analysis presented in this work is aimed at determining fundamental QCD (Quantum Chromodynamics) parameters, the BaBar physics program includes the making of many important physics measurements, including ones in other $B, \tau$, charm and two-photon processes ${ }^{1}$. Because CP asymmetries in the $B$ system are expected to be large, measurements with an accuracy of $10 \%$ can be made with several hundred reconstructed events, however, typical branching ratios are on the order of $10^{-5}$. Thus tens of millions of $B$ pairs must be produced to yield an appreciable number of $B^{\prime}$ s decaying in these modes. To deliver this many $B^{\prime}$ s to the BaBar detector, the design luminosity of PEP-II was $3 \times 10^{33} \mathrm{~cm}^{-2} s^{-1}$ which yielded $\sim 3 B$ pairs per second being produced. This

[^3]design luminosity was surpassed in 2001 and the current typical daily luminosity in the summer of 2004 is $\sim 8 \times 10^{33} \mathrm{~cm}^{-2} \mathrm{~s}^{-1}$.


Figure 4.1: BaBar detector; the standard coordinate system is defined with $+z$ pointing along the electron beam direction (to the right in this figure), and $\theta$ and $\phi$ are then the standard polar and azimuthal angles with respect to this axis.

The BaBar detector is composed of five major coaxial subdetectors, which are, listed from closest to the beampipe to farthest from it:

- The Silicon Vertex Tracker (SVT)
- The Drift Chamber (DCH)
- The Detector of Internally Reflected Cerenkov Light (DIRC)
- The Electromagnetic Calorimeter (EMC)
- The Instrumented Flux Return (IFR)
and a super-conducting solenoid located between the DIRC and EMC which provides a uniform coaxial 1.5 T magnetic field.

The subdetectors are generally cylindrically shaped in the central or "barrel" region to detect particles which come out relatively transverse to the beam axis, but the EMC and IFR also have active endcap regions to detect those particles which are propagating more parallel to the beam.

### 4.2 The Positron-Electron Project (PEP) II



Figure 4.2: SLAC linac and storage rings.

The SLAC Linac is a two mile long linear accelerator which provides high energy electrons and positrons to the PEP-II storage ring. The linac was completed in 1966 and has seen several upgrades and changes over its 38 year history. The Positron-Electron

Project-II (PEP-II) storage ring is also an upgraded machine. It started out as PEP, a 30 $\mathrm{GeV} / c$ storage ring used through the 1980's to test the theory of QCD. The PEP tunnel is $\sim 800$ meters in diameter and currently houses the PEP-II storage ring. The most striking feature of the current configuration of the SLAC linac is that it delivers the electrons and positrons with different beam energies into the PEP-II storage ring.

### 4.2.1 Asymmetric Collider

The asymmetric mode of operation is a novel solution to a very challenging problem. The problem is that when the $\Upsilon(4 \mathrm{~s})$ (mass $=10.58 \mathrm{GeV} / c^{2}$ ) decays into two $B^{0}$ 's (mass $=5.28 \mathrm{GeV} / c^{2}$ each), there is very little energy left over in the form of kinetic energy. Each $B^{0}$ will propagate along with a momentum of $\sim 300 \mathrm{MeV} / c$ in the $\Upsilon(4 \mathrm{~s})$ rest frame. Thus a typical $B^{0}$ will only travel about $26 \mu \mathrm{~m}$ before it decays. This is too small for current vertex trackers to accurately resolve. However, by colliding asymmetric beams, the $\Upsilon(4 \mathrm{~s})$ is given a large boost which increases both the velocity and time dilation for the decaying $B^{0 \prime}$ s. In BaBar, $9.0 \mathrm{GeV} / c$ electrons are collided with $3.1 \mathrm{GeV} / c$ positrons to produce a $\Upsilon(4 \mathrm{~s})$ with a lab-frame energy of 12.1 GeV and $\beta \gamma=0.56$. This produces an average $B^{0}$ separation of $\sim 250 \mu \mathrm{~m}$, still challenging, but much more manageable from a resolution point of view.

### 4.2.2 Accelerating Particles

Accelerating electrons and positrons to their final energies is accomplished in several steps. Electrons are first produced by an electron gun - a device similar to the
electron gun at the back of a television set, but with significantly higher capacity. In the case of a TV, a filament is heated and an adequate number electrons are released to form the picture on the screen. In the case of the PEP linac though, a laser is fired into a metal target ablating much larger numbers of electrons into an area with a strong electric field. This field then accelerates the electrons up to $\sim 10 \mathrm{MeV} / c$; at the same time positrons are constantly being produced by siphoning off some of the high energy electrons from the linac and colliding them with a fixed tungsten target which produces electron-positron pairs. The positrons are selected out by a magnetic field, collected in bunches and returned to the start of the linac where they follow the same path as electrons to the damping rings.

The reason for this is that a narrowly focused beam of particles will produce more collisions than a diffuse beam, so the goal of the damping ring is to dissipate motion not in the beam direction. However, both beams at this point are still relatively spread out, so the electron (positron) beam is sent to the north (south) damping ring. Each time an electron or positron completes a cycle in the damping ring it loses energy to synchrotron radiation and receives a boost equal to the synchrotron loss in the beam direction. Synchrotron radiation dissipates energy in the direction that the particle is moving, so particles with momentum transverse to the beam direction will dissipate some of this transverse momentum. The particle will then receive a boost in the beam direction and the net effect is a focussing of the beam in the beam direction.

The next step is to return the electrons and positrons to the linac where they will
be accelerated to their full energy.
The energy the particles gain in the linac is proportional to how far they travel in the linac. The lower energy positrons are removed first, then the higher energy electrons (as shown in Figure 4.2). Both beams are kicked out of the linac in bunches and sent to the PEP-II storage ring in opposite directions, where the circulate in counter-rotating directions and are ultimately steered together to collide at the collision point in the center of the BaBar detector ${ }^{2}$.

### 4.3 The Silicon Vertex Tracker (SVT)



Figure 4.3: SVT profile; note five-layer concentric shell structure.

The innermost subdetector in Figure 4.1 is the Silicon Vertex Tracker or SVT. The SVT is composed of 5 double sided silicon strip detector layers, as shown in Figure 4.3. The inner sides of these layers have strips oriented perpendicular to the beam axis to

[^4]measure $z$ position (along the beam axis) while the outer surfaces have orthogonal strips to measure the $\phi$ coordinate. The SVT geometry was determined by a need to accommodate the final beam bending magnets, located at $\pm 20 \mathrm{~cm}$ from the central interaction point, and to cover the largest amount of solid angle (measured relative to the interaction point as the center) as possible. The first three layers are flat modules which are critical for accurately measuring the $B^{0}$ vertices. The fourth and fifth layers have an arched design which maximizes solid angle coverage while avoiding large track incidence angles. Information from them added to that of the inner layers helps to determine $B^{0}$ vertices, serves to align SVT tracks with Drift Chamber (DCH) tracks and is important for tracking charged particles which have a transverse momentum of less than $100 \mathrm{MeV} / c$, since these particles will not enter the DCH. The internal structure of the silicon strips within the sublayers of the SVT is detailed in Table 4.1.

During construction and material choice, particular attention is paid to the radiation hardness of the SVT components as the SVT accumulates a very large radiation dose due to its proximity to the beam pipe. At the same time, an attempt is made to minimize the amount of material in the SVT to minimize the effect of multiple scattering of particles (which leads to resolution-induced smearing of the true initial track direction) as they pass through the body of the SVT.

The SVT positional resolution is shown in Figure 4.4 in both $z$ and $\phi$ as a function of incident track angle. In fact, SVT resolution is dominated by multiple scattering, not SVT construction or electronics hardware.


Figure 4.4: SVT resolution for data and MC as a function of incident angle.

Because of the beam energy asymmetry, it is extremely important to maximize coverage in the forward direction of the SVT (and entire detector), since the asymmetric boost will increase the number of decay tracks that will travel that direction. For this reason, all instrumentation readout and cooling components are located in the backward support cone of the SVT (the support cone is the piece of the SVT upon which all the SVT pieces are mounted, and which in turn is mounted directly upon the beampipe). This can be seen by the asymmetry between the shape of the forward and backward support cones in Figure 4.3.

The SVT is very significant for reconstruction of the $D^{*} \ell \nu$ decay because many

| Layer | Radius <br> $(\mathrm{cm})$ | Modules | $\phi$ pitch <br> $(\mu \mathrm{m})$ | z pitch <br> $(\mu \mathrm{m})$ |
| :---: | :---: | :---: | :---: | :---: |
| 1 | 3.2 | 6 | 50 or 100 | 100 |
| 2 | 4.0 | 6 | 55 or 110 | 100 |
| 3 | 5.4 | 6 | 55 or 110 | 100 |
| 4 a | 12.4 | 8 | 100 | 210 |
| 4 b | 12.7 | 8 | 100 | 210 |
| 5 a | 14.0 | 9 | 100 | 210 |
| 5 b | 14.4 | 9 | 100 | 210 |

Table 4.1: SVT sublayers, radius and pitch (distance between each of the active silicon detection strips in the given sublayer).
of the intrinsically low momentum pions from the direct decay of the $D^{*}$ do not reach the next subdetector out (the DCH), and thus are only reconstructed from SVT information.

### 4.4 The Drift Chamber (DCH)

The Drift Chamber or DCH is the primary tracking device for BaBar. It provides a precision measurement of transverse momentum $\left(p_{T}\right)$ for charged tracks with $p_{T}$ greater than $120 \mathrm{MeV} / c$ (particles with momenta below this threshold generally do not have enough DCH hits to form tracks that are initially reconstructed with the DCH track-finding software). The DCH starts at a radius of 23.6 cm and extends out to 79.0 cm . The DCH contains 7104 cells, each approximately 1.2 cm by 1.8 cm . The 40 layers are organized into 10 superlayers, each four wire superlayer has a uniform angle alignment for all of its sublayers. Four of the superlayers (1, 4, 7, and 10) are axial (A) superlayers, which means that the wires run perfectly parallel to the detector $z$ axis. The remaining superlayers have a non-zero stereo angle which allows us to determine the $z$ position as
well as radius and $\phi$ of a measured track. The stereo angle ranges from 45 mrad in superlayer 2 out to 76 mrad in super-layer 9 and the sign of the angle alternates every other stereo layer creating two different types of stereo superlayers ( U - positive stereo angle, V - negative stereo angle).

A very schematic sideview of the DCH is shown in Figure 4.5 (indicating only dimensions but not the angles of the wires in the stereo superlayers, here).


Figure 4.5: Side view of the Drift Chamber layout, the high energy electrons are traveling to the right. As with the SVT, the interaction point (IP) is located closer to the high energy side to improve solid angle coverage as the particles spray preferentially out towards the right hand (low energy) side. Note dimensions are given in millimeters in this diagram.

Each of the 7104 DCH cells consists of a single central sensing wire, maintained at a voltage of 1900V-1960V (depending on the period of running), and six field shaping wires. For cells on the inner side of a superlayer, the field shaping wires are grounded on the rear end plate. For cells on the outer side, one wire is held at 350 V . The positional resolution for each cell is shown in Figure 4.6, $125 \mu \mathrm{~m}$ is the average cell resolution.

The DCH uses a mixture of He:Isobutane ( $80 \%: 20 \%$ ) as its counting gas. This

| Superlayer | Cells/layer | Inner radius (cm) | Layer Type |
| :---: | :---: | :---: | :---: |
| 1 | 96 | 26.0 | A |
| 2 | 112 | 31.9 | U |
| 3 | 128 | 37.1 | V |
| 4 | 144 | 42.3 | A |
| 5 | 176 | 48.1 | U |
| 6 | 192 | 53.3 | V |
| 7 | 208 | 58.5 | A |
| 8 | 224 | 64.3 | U |
| 9 | 240 | 69.5 | V |
| 10 | 256 | 74.7 | A |

Table 4.2: DCH superlayers, layer type is axial(A) or stereo(U,V).
mixture has a low density (to minimize multiple scattering for good $p_{T}$ resolution), good spatial and $\mathrm{dE} / \mathrm{dx}$ resolution, and reasonably short drift time (time for the ionized electrons to drift to the signal detection wires). The gas and wires total to $0.3 \%$ radiation lengths $\left(X_{0}\right)$. The inner support cylinder ( 1 mm beryllium) contributes $0.28 \% X_{0}$ while the outer carbon fiber support cylinder contributes $1.5 \% X_{0}$.

The DCH serves at least two primary purposes other than tracking. First, the DCH provides prompt information to the Level 1 Trigger system (see Section 4.8) which is a coarse filter for removing unwanted events and saving valuable processing time. And second, the DCH also has a crucial role in particle identification (PID). Figure 4.7 shows how the energy loss of a particle per unit distance ( $\mathrm{dE} / \mathrm{dx}$ ) vs. momentum correlates to a particular particle type. This information can be combined with information from other subdetectors to estimate the PID of charged tracks in each event.


Figure 4.6: Drift Chamber resolution vs track distance from wire at 1961V.

### 4.5 The Detector of Internally Reflected Cerenkov Light (DIRC)

If particles have a transverse momentum of greater than $\sim 180 \mathrm{MeV} / c$, they will be able to make it past the outer layer of the DCH without curling up inside the DCH volume, and they will then enter the Detector of Internally Reflected Cerenkov Light or DIRC. Partly due to space constraints within the detector, the BaBar DIRC design is rather novel compared with previous Cerenkov light detectors for collider experiments (which are usually gas-filled), and we describe its relevant design and functional principles in this section.

The DIRC consists of 144 amorphous fused silica bars (which are colloquially simply referred to as "quartz" bars) arranged into 12 "bar boxes". Each bar is 1.7 cm thick, 3.5 cm wide and 490 cm long. Charged particles traveling through a bar will radi-


Figure 4.7: $\mathrm{dE} / \mathrm{dx}$ vs p from Drift Chamber; the units on the vertical axis are proportional to energy (loss) per unit length.
ate Cerenkov light whenever the particle is traveling above the speed of light in quartz (which is equal to c/(index of refraction of quartz), or about $2 / 3$ the speed of light in vacuum). The Cerenkov light is guided to the instrumented section of the DIRC via total internal reflections with the quartz bar side faces. The uninstrumented end of the quartz bar is mirrored to return photons back to the instrumented end. The general structure of the DIRC is shown in Figure 4.8.

The instrumented end then connects to a large, half toroidal tank (the "Standoff Box") filled with purified water which closely matches the index of refraction of the quartz bar to minimize transmission losses. The angle between the light cone and the


Figure 4.8: DIRC structure, emphasizing quartz bars, support framework, and Standoff Box.
emitting particle is preserved by the internal reflections. Once the light arrives at the water tank, the individual photons travel through the water until they are detected by one of the $\sim 11,000$ Photomultiplier Tubes (PMTs). This principle is shown in Figure 4.9. The large Standoff Box radius allows for a finer resolution of the Cerenkov angle. The resolution for a single photon is 10.2 mr , as shown in Figure 4.10. There are generally $\sim 30$ photons detected each event so by using pattern recognition, the DIRC can improve angular resolution down to about 2.8 mr . This resolution allows for a 3 sigma separation between pions and kaons at about $3 \mathrm{GeV} / c$. The DIRC is extremely critical for good particle identification (especially for hadrons).


Figure 4.9: DIRC operating principle, showing total internal reflection of the photons, and their detection in PMT's in the Standoff Box.

### 4.6 The Electromagnetic Calorimeter (EMC)

After passing through the DIRC bars, particles will enter the Electromagnetic Calorimeter or EMC, which is constructed with a more standard design.

The BaBar EMC is composed of 6580 cesium iodide crystals. 5760 of these crystals are contained in the barrel region, which is divided into 48 polar-angle rows ( $\theta$ direction) with 120 crystals in each row ( $\phi$ direction). The forward endcap region contains the remaining crystals in eight polar-angle rows. Each crystal is a trapezoidal pyramid with a front face surface area of $\sim 5 \mathrm{~cm} \times 5 \mathrm{~cm}$ and a depth of $29.76-32.55 \mathrm{~cm}$, with the crystal length increasing in the forward direction (the increase corresponds to an increase from 16 to 17.5 radiation lengths). All cooling, electronics and almost all support material is


Figure 4.10: DIRC single photon resolution and photons per track.
located behind the crystals to minimize the material in front of the crystals.
The EMC functions by either absorbing photons or by the cesium iodide in the crystals interacting electromagnetically with charged particles from an event and subsequently emitting scintillation light. The scintillation light is detected by two photodiodes on the back of each of the crystals and converted into digital signals. The photon energy resolution and angular resolution at 90 degrees are shown in Figure 4.11.

An important issue for the calorimeter is calibration. The energy detection range extends from $\sim 10 \mathrm{MeV} / c$ up to $9 \mathrm{GeV} / c$ and is calibrated in four separate, overlapping regions. Several methods of calibration are used:

- Injection of charge: Charge is injected at the preamplifier input, with predetermined magnitude, phase and pattern. This tests in detail the response of each amplifier. This calibration takes several minutes.
- Bhabha events: $e^{+} e^{-}$pairs can be used for calibration, with the running time for


Figure 4.11: EMC position and energy resolution (the $\sigma^{\prime}$ s are the calibration-derived parameters that determine the resolution of the detector).
calibration being less than one day of normal running with collisions. This calibration data is taken simultaneously with normal data taking.

- Liquid radioactive source: $6.1 \mathrm{MeV} / \mathrm{c}$ photons from ${ }^{16} N$ decays are detected and measured, crystal response over time recorded and monitored.
- Light pulser: Turned on when needed, convenient for tracking short term changes in crystal response.

Calibration helps track short and long term changes in the crystal response. Crystal response can vary due to radiation damage or damage to optical surfaces and couplings. Calibration also measures constants used for maximizing the resolution of the EMC.

### 4.7 The Instrumented Flux Return (IFR)

High energy particles that make it all the way through the EMC and the magnet solenoid will enter the Instrumented Flux Return or IFR.

The IFR serves both as a flux return for the 1.5 T solenoidal magnet and as a muon and neutral hadron detector. The IFR has three main segments: the barrel and forward and backward endcaps. The barrel region extends from a radius of 1.88 m to 3.23 m and is divided into sextants. The endcaps are hexagonal plates divided into two halves to allow the IFR to be moved into place around the inner subdetectors. The IFR consists of 18 iron plates which vary in thickness. Interleaved between these plates are active detector layers, the Resistive Plate Chambers (RPC's). One double layer cylindrical RPC resides between the EMC and IFR, 17 layers exist between the 18 IFR iron plates and two more active layers are located just outside the last IFR plate. There are 18 active layers in the endcaps (one prior to the first IFR plate plus 17 in between plates).

Detailed Monte Carlo studies showed that low momentum muon and $K_{L}^{0}$ detection improve with thinner absorbing plates, but this effect is only significant for the first absorption length. As a result, the first nine layers of the IFR are 2 cm thick, then four are 3 cm thick, followed by three at 5 cm and two at 10 cm . In the endcaps, one of the 10 cm plates is replaced with a 5 cm plate.

RPC's consist of two 2 mm thick Bakelite plates separated by 2 mm , with outer surfaces painted with graphite of high surface resistivity ( $\sim 100 \mathrm{k} \Omega$ /square) and covered with an insulating film. One graphite surface is connected to high voltage ( $\sim 8 \mathrm{kV}$ ) while
the other is connected to ground. The inner surfaces (the two Bakelite surfaces which face each other) are treated with linseed oil to enhance high efficiency and to attain low noise. Each RPC is 3.2 cm tall, 125 cm wide and 181 cm to 320 cm long (depending on the location of RPC). The active space between the two plates is filled with a gas mixture of argon, Freon and a very small amount ( $\sim 5 \%$ ) of ISO-butane. This gas mixture is non-flammable and environmentally safe.

The RPC operates as follows: when a charged particle crosses the chamber a quenched spark is produced in the gas which produces an electrical signal, which is then read out on the backside of the chamber.

### 4.8 The Trigger System

The BaBar Trigger is not a separate subdetector, but a system that takes information from all the subdetectors and makes critical decisions with it. The Trigger is composed of the Level 1 Trigger (L1 Trigger) and the Level 3 Trigger (L3 Trigger). The task of the Trigger system is to very rapidly filter out uninteresting events while maintaining nearly $100 \%$ efficiency for $B$ physics events.

The L1 Trigger is a dedicated hardware-based filtering system which has a goal of coarsely reducing millions of events per second down to less than 2000 which can then be passed on to the L3 Trigger, which then has much more time to filter through the events and make a decision about whether or not to keep the event. If an event passes the L3 Trigger it is written to tape. The passthrough goal for the L3 Trigger is about 100
events per second.

### 4.8.1 Level 1 Trigger

The L1 Trigger is comprised of the Drift Chamber Trigger (DCT), the Electromagnetic Calorimeter Trigger (EMT), the Instrumented Flux Trigger (IFT) and the Global Level 1 Trigger (GLT). The DCT produces three 16-bit maps of $\phi$ coordinates for candidate tracks which pass selection as either a short track ( $p_{T}>120 \mathrm{MeV} / c$ track which reaches Superlayer 5 of the DCH), a long track ( $p_{T}>180 \mathrm{MeV} / c$ track which reaches Superlayer 10 of the DCH ) or a high momentum track $\left(p_{T}>800 \mathrm{MeV} / c\right)$. The EMT produces five 20-bit $\phi$ maps of particle candidates depending on the amount of energy deposited and location, see Table 4.3. These maps plus a single $\phi$ map from the IFT are collected by the GLT to form 24 trigger lines which are passed to the L3 trigger, see Figure 4.12.

| EMT objects |  | DCT objects |  |  |  |
| :--- | :--- | ---: | :--- | :--- | ---: | :--- |
|  | Cluster Description | Energy Cut |  | Description | $P_{t}$ Cut |
| M | Minimum ionizing | $100 \mathrm{MeV} / c$ | B | Track reaching SL-5 | $120 \mathrm{MeV} / c$ |
| G | Intermediate ionizing | $300 \mathrm{MeV} / c$ | A | Track reaching SL-10 | $180 \mathrm{MeV} / c$ |
| E | High energy e/ $\gamma$ | $700 \mathrm{MeV} / c$ | A $^{\prime}$ | High $p_{t}$ track | $800 \mathrm{MeV} / c$ |
| X | forward endcap | $100 \mathrm{MeV} / c$ |  | reaching SL-10 |  |
| Y | Backward barrel | $1 \mathrm{GeV} / c$ |  |  |  |

Table 4.3: Trigger $\phi$ map objects defined for the EMT and DCT. The IFT has a single $\phi$ map labelled U which indicates a track in the IFR (SL = DCH SuperLayer).

The L1 Trigger system was designed to have two highly efficient orthogonal internal triggers, the DCT and EMT. Most of the 24 Trigger lines (13/24) are purely DCT or EMT triggers. This allows the trigger efficiencies to be easily cross checked and mea-



Figure 4.12: Left: Elements which make up the Level 1 trigger. The TSF (Track Segment Finder) sends track segment information to the BLT (Binary Link Tracker) and PTD ( $P_{t}$ Discriminator). The BLT, PTD, EMT and IFT all send their respective bit map information to the GLT which decides if a Level 1 accept should be issued. Right: DCT efficiency plot. The A and B tracks turn on curves are determined by the minimum $p_{t}$ required to reach half way and all the way through the Drift Chamber. The A' track turn on is programmable, currently set at $800 \mathrm{MeV} / c$.
sured. For $B^{0} \bar{B}^{0}$ events, each trigger is $>99 \%$ efficient alone and $>99.9 \%$ combined. Thus, nearly every single $B^{0} \bar{B}^{0}$ event is triggered on, captured, and written to tape.

After the GLT collects all the L1 Trigger information, it decides whether to issue a L1 Trigger accept or not. If a L1 accept signal is sent by the GLT to the other trigger boards, the event data currently stored in memory is passed to the L3 Trigger, if no L1 accept is received, the data is overwritten and lost. A L1 accept decision sends L1 information to L3 within 11-12 $\mu \mathrm{s}$ after the corresponding event occurred.

### 4.8.2 Level 3 Trigger

The L3 Trigger (L3T) is a software based filter which takes the greatly reduced number of events from the L1 Trigger and applies further constraints on selecting good events to be stored on disk for future use. The L3T runs in parallel on 32 computer nodes. The L3T takes data from the DCH and EMC and forms track and cluster objects. These objects are passed through a series of filters designed to eliminate backgrounds such as tracks not coming from the interaction region. Events which pass the L3T filters are written to disk to be processed later. Figure 4.13 shows the L3T event display. This display shows the raw tracks (red lines), EMC clusters (red towers) and activated trigger lines ( $2 \mathrm{E}, \mathrm{EM}^{*}$, etc...).

The L3 trigger creates composite objects from the fundamental EMT and DCT objects. The trigger line definitions change over time, Figure 4.13 shows trigger lines used on Oct 28, 2000. The naming conventions used by the L3 trigger are shown in Table 4.4.

Once an event passes both L1T and L3T and is written to tape, it is ready to be processed in the next step: offline analysis.


Figure 4.13: L3 event display. The blue circles are track segments from the TSF boards. The red lines label reconstructed charged tracks and the red towers show energy deposits reported by the EMT.

| Back-to-Back objects |  |  |
| :---: | :---: | :---: |
| Object | Description | $\phi$ Cut |
| $B^{*}, A^{*}$ | back to back short, long tracks | $124^{\circ}$ |
| $M^{*}, G^{*}$ | back to back M, G clusters | $117^{\circ}$ |
| E-M | $E$ and $M$ clusters back to back | $126^{\circ}$ |
| DCT+EMT matching objects |  |  |
| BM | $B$ and $M$ with matching phi | $<27^{\circ}$ |
| AM | A and M with matching phi | $<27^{\circ}$ |
| $\mathrm{A}^{\prime} \mathrm{M}$ | $\mathrm{A}^{\prime}$ and M with matching phi | $<27^{\circ}$ |
| Compound Trigger objects |  |  |
| A+ | $1 \mathrm{~A} \& \mathrm{~A}^{\prime}$ |  |
| D2 | 2B \& 1A |  |
| D2* | $B^{*} \& 1 \mathrm{~A}$ |  |
| $D 2^{*}+$ | $B^{*} \& 1 \mathrm{~A}+$ |  |
| $Z^{*}$ | 2 Z with loose back-to-back cut, Z is any primitive |  |

Table 4.4: Trigger objects defined for the EMT and DCT. The IFT has a single object labelled U which indicates one or more tracks in the IFR.

## Chapter 5

## Event Selection Description for <br> $B^{0} \rightarrow D^{*} \ell \nu \quad$ Candidates

In this Chapter we will give a summary of the relevant steps of our event selection, and give a description of our MC simulation framework.

### 5.1 BABAR Monte Carlo (MC) Description

As we will be showing results from BaBar MC in the following sections (and throughout the rest of this work), it is useful to initially describe some details of the BaBar MC simulation system.

### 5.1.1 Generator Level MC - EvtGen

A program called EvtGen forms the kernel of the BaBar MC production system. Events are generated with fully known initial beam momenta (for the incoming $e^{+}$and $e^{-}$), and user inputs for the branching ratios to each decay mode. The four-momenta of all outgoing particles are determined precisely, and there is no smearing involved, nor any detector simulation whatsoever - these all come later as the EvtGen tracks are propagated through the detector, as part of the full BaBar MC simulation. All input parameters (such as the form factors) for the event generation are controlled, so this is a useful minilaboratory to test various features of the decaying system (such as the shape dependence of the kinematic variables as a function of the form factors, as we will see later). However, no realistic acceptance or smearing studies can be done on EvtGen samples, since they include no detector simulation.

Validation of EvtGen MC production vs. our fitting tools (RooFit) is shown in Appendix C.

### 5.1.2 Full MC - Simulation Production 4

SP4 (Simulation Production iteration \#4) is the fourth iteration of the current cycle of the full BaBar MC , which exercises the full BaBar detector simulation and software reconstruction chain. It begins with the EvtGen software package, described in Section 5.1.1. EvtGen forms the kernel event and particle generator for all BaBar MC production. From here, SP4 uses the GEANT4 package to build a total simulation of each
of the primary five BaBar subdetectors. This simulated detector then takes as inputs the various charged and neutral particles and directions that EvtGen has determined for a particular event. These are then swum through the detector simulation, and hits and clusters are deposited in various subdetectors. These hits are then reconstructed into individual tracks and neutral candidates, which the reconstruct software ultimately builds into a single event.

In the case of our analysis we take this candidate $D^{*} \ell \nu$ event, and then as with data, impose further cuts at various stages to distinguish and classify this event between one that is most likely signal, and others that are more likely background.

SP4 used version 10 of the BaBar software framework (i.e., Release 10 processing), which was the same software version used initially to process all data from Runs $1+2,2000-2002$. This is the version of the processed data that we use, so that we have consistency between MC and data software reconstruction versions.

SP3 was the previous iteration of MC and was used in early stages of this analysis when it corresponded to the the version of the software being used for the data reconstruction as well. The primary differences between SP3 and SP4 were in the detector model, which was much improved in the latter, and though some of the discussion in places in this work refers to earlier tests done with SP3, all our ultimate results were done only on SP4. However, since some of our earlier tables and figures were made with SP3, we have left them in this work for comparison purposes, and labelled them clearly when they were made with this earlier MC iteration.

## $5.2 \quad B^{0} \rightarrow D^{*} \ell \nu$ Event Selection

Now that we have seen a description of the BaBar MC system, and the relevant components of the BaBar detector in Chapter 4, we turn our focus on how we reconstruct a candidate $B^{0} \rightarrow D^{*} \ell \nu$ event from the various types of energy deposits left in the detector.

After this, we summarize our final selection cuts.

### 5.2.1 Initial Data/MC Processing

We take for our analysis as inital input the common Data and MC ntuples used also for the BaBar branching ratio and $\left|V_{c b}\right|$ measurements. These were made using the DstarlnuUser package (see Section C. 3 for more information on this package).

Our group further processed the common ntuples into "rtuples" (data format readable by Root) and further refined and consolidated multiple rtuples into single large rtuples for chunks of MC or data so that these sources could then be compared and analyzed by macros within the RooFit framework (we will describe later the final cuts made at this level).

### 5.2.2 Decay Channels Considered

This analysis determines the three form factors involved in neutral $B$ meson semileptonic ( $B^{0} \rightarrow D^{*} \ell \nu$ ) decays. Because the neutrinos from this decay escape detection and measurement, we are missing a powerful constraint which analysts of the
hadronic $B^{0}$ decay channels can utilize to ensure that all the decay daughters add up to a $B^{0}$. As a result, the $D^{*} \ell \nu$ channel requires very careful analysis of backgrounds and errors.

The full decay chain we use in this analysis is $B^{0} \rightarrow D^{*} e \nu, D^{*} \rightarrow D^{0} \pi$, and $D^{0} \rightarrow K^{-} \pi^{+}$.

Earlier, work was done on three other decay modes of the $D^{0}: K \pi^{+} \pi^{0}, K^{-} \pi^{+} \pi^{+} \pi^{-}$, and $K_{S} \pi \pi$. For larger statistics, future results will potentially include the former two modes, however the events are relatively few and the backgrounds quite high in the last mode so inclusion of this mode is not foreseen in the near future (particularly as higher statistics will soon not be the limiting aspect on the total error for this analysis). However, some of our earlier tables and figures still include this mode, as well as the other three modes, and we have left them in for comparison purposes as well.

The $K^{-} \pi^{+}$mode is by far the cleanest mode in terms of signal to background ratio of all the $D^{0}$ decays (the $K \pi^{+} \pi^{0}$ and $K^{-} \pi^{+} \pi^{+} \pi^{-}$modes have larger branching ratios but also larger backgrounds due to higher combinatorics, and difficulty in reconstructing the correct $\pi^{0}$ with high photon backgrounds).

Even in the $K^{-} \pi^{+}$mode with electrons only, we already have nearly 17 times the number of events as in the previous CLEO analysis for the 2000-2002 BABAR dataset alone.

### 5.2.3 Data and MC Set

The data used for this analysis comes from "good runs" starting in 2000 and ending in 2002. It correponds to $79 \mathrm{fb}^{-1}$ collected on the $\Upsilon(4 s)$ resonance which produced approximately $85 \times 10^{6} B \bar{B}$-pairs. There are $\approx 8.6 \times 10^{6} B^{0} \rightarrow D^{*} \ell \nu$ decays in this sample of which we have reconstructed 16,061 using only the electron channel and the decay mode $D^{0} \rightarrow K^{-} \pi^{+}$.

About three times as much equivalent luminosity generic SP4 Monte Carlo data was processed and used for validation studies and calibration, as well as a similar amount of $D^{0} \rightarrow K^{-} \pi^{+}$mode signal MC. All available SP4 $B^{0}, B^{+}, c \bar{c}$ and signal Monte Carlo that are reconstructed with Release 10 were processed. The exact number of events generated in each mode for each year are shown below:

```
2001:
nB0B0:58,584,153
nBpBm:62,342,000
nCCbar:92,584,401
2002:
nB0B0:35,714,948
nBpBm:34,371,142
nCCbar:27,615,188
2000:
nB0B0:29,461,186
nBpBm:30,748,192
nCCbar:41,332,550
```

We weight the number of charged $B$ and $c c b a r$ events to the number of neutral $B^{\prime} \mathrm{s}$, assuming that the cross-section at the $\Upsilon(4 \mathrm{~S})$ to charged and neutral $B^{\prime}$ 's is the same, and that the cross-section to $c c b a r$ events is in the ratio $\frac{1.3}{1.1}$ (since the cross-section at the
$\Upsilon(4 \mathrm{~S})$ peak for $e^{+} e^{-} \rightarrow c \bar{c}$ is 1.3 nb , and for $e^{+} e^{-} \rightarrow b \bar{b}$ is 1.1 nb$)$.
Adding up the number of $B$ events and dividing cross-section we find that the equivalent luminosity that these events correspond to is:

$$
\begin{equation*}
(2 \times 58.6+2 \times 35.7+2 \times 29.5) * 10^{6} / 1.1 n b \approx 225 f^{-1} \tag{5.1}
\end{equation*}
$$

### 5.2.4 Backgrounds

Some of the most challenging aspects of this measurement are the complications due to various backgrounds to the reconstruction of a $B^{0} \rightarrow D^{*} \ell \nu$ candidate. The reconstruction is essentially a three step process. First the $D^{0}$ candidate is found in one of the four decay modes listed in Section 5.2.2. Then a slow pion candidate is added to the $D^{0}$ to create a $D^{*}$ candidate and finally a lepton candidate is added to the $D^{*}$ to create the $B^{0} \rightarrow D^{*} \ell \nu \quad$ candidate. The mass difference between the $D^{*}$ and $D^{0}, \delta m$, is just larger than the charged $\pi$ mass and thus forms a very narrow peak when plotted for reconstructed signal events. The combinatoric background forms a very broad sideband in $\delta m$, so this peak serves as a very powerful cut against backgrounds coming from misreconstructed $D^{0}$ or $D^{*}$ candidates (see Figure 5.19 for a typical distribution).

Unfortunately, there is no such constraint for the final leg of the reconstruction. The neutrino escapes detection and the ability to add up all the daughter energies to a total whose four-momentum is equal to that of the parent is lost. This manifests itself in the analysis as the unavoidable ultimate presence of numerous backgrounds which involve a good $D^{*}$ but a bad lepton candidate in the final step of reconstructing the $B^{0} \rightarrow D^{*} \ell \nu$
candidate.
We may categorize the main backgrounds into six types for this analysis:

- Combinatoric background: The $D^{0}$ is incorrectly reconstructed or the slow pion candidate is, or both of them are, or either or both do not come from a real $D^{*}$. Cutting on $D^{*}-D^{0}$ mass $(\delta m)$ is effective at removing these backgrounds.
- Fake lepton: A real $D^{*}$ is combined with a hadron instead of a lepton. This background is minimized by requiring very tight lepton identification. Lepton fake rates are well measured.
- Correlated lepton: A good $D^{*}$ combined with a secondary lepton from the same $B^{0}$ or $B^{-}$parent. $B^{0}$ or $B^{-} \rightarrow D^{*} X, X \rightarrow l \mathrm{Y}\left(\mathrm{X}=\right.$ e.g. another $D^{*}$ or $\left.D^{0}\right)$. Secondary leptons have a softer momentum spectrum than primary leptons so that the lepton momentum cut is effective against this background.
- Uncorrelated lepton: A good $D^{*}$ combined with lepton from opposite side B. $B^{0}$ or $B^{-} \rightarrow D^{*} X, \overline{B^{0}}$ or $B^{+} \rightarrow Y l$. Several angular cuts are effective and help define a control sample for estimating the size of this background.
- $D^{* *}$ background: A charged $B$ can decay into a $D^{* *}$ which quickly decays into a $D^{*}$ plus at least one slow pion. $B^{ \pm} \rightarrow D^{*} l \nu+n \pi_{\text {slow }}$. Though the $\cos \theta_{B Y}$ cut removes some of this background, it is is difficult to tell this category apart from true signal events and fully remove it, and some of it will remain in the final sample after all cuts.
- Continuum: A $c \bar{c}$ event can have one $c$ quark form a $D^{*}$ while the other hadronizes into a state which decays semileptonically to create a lepton $\left(c \bar{c} \rightarrow D^{*} l X\right.$; the other continuum backgrounds ( $u d s$ ) are completely negligible and generally almost none of them pass the final cuts). This background is very jetty, cuts on the $D^{*}$ momentum and event topology are effective at suppressing it.


### 5.2.5 Reconstruction track lists

The particles that are actually observed in the BaBar detector are pions, electrons, muons, kaons, protons and photons. All other particles are reconstructed from these basic building blocks. These particles are detected as charged tracks or neutral clusters, and they are grouped into the following lists depending on the quality and characteristics of the track/cluster [26]:

- ChargedTracks: All reconstructed tracks from DCH or SVT. The pion mass hypothesis is used for the $\mathrm{d} E / \mathrm{d} x$ deposition to reconstruct the tracks.
- GoodTracksVeryLoose: subset of ChargedTracks, must pass:

1. a distance of closest approach to the per-event beam spot of $|\Delta z|<10 \mathrm{~cm}$, and $\sqrt{\Delta x^{2}+\Delta y^{2}}<1.5 \mathrm{~cm}$
2. a maximum track momentum measured in the lab frame of 10 Gev
3. a minimum number of DCH + SVT track hits $\geq 5$

- GoodTracksLoose: a subset of GoodTracksVeryLoose, must pass:

1. a minimum transverse momentum of $100 \mathrm{MeV} / c$
2. a minimum number of 12 track hits recorded in the DCH (out of a maximum of 40 possible).

- GoodTracksTight: a subset of GoodTracksLoose, must pass:

1. a distance of closest approach to the per-event beam spot of $|\Delta z|<3 \mathrm{~cm}$, and $\sqrt{\Delta x^{2}+\Delta y^{2}}<1 \mathrm{~cm}$
2. a minimum number of 20 track hits recorded in the DCH

- GoodPhotonLoose

1. a minimum calorimeter energy of 30 MeV
2. a minimum number of EMC crystals hit $>0$
3. LAT $<0.8$ (LAT is an energy deposit shape variable, larger for hadrons than for photons and electrons, see Section 5.2.7.2)

- GoodNeutralLoose

1. a minimum calorimeter energy of 30 MeV
2. a minimum number of EMC crystals hit $>0$
3. no LAT cut

These lists serve as a pool of candidates for creating composite particles such as the $D^{0}$. See Figure 5.1 for a comparison of several basic track-level quantities between these various categories of tracks.


Figure 5.1: Distributions of (a) total momentum, (b) transverse momentum, (c) cosine of polar angle, and (d) azimuthal angle (all measured in the lab frame) for the following lists: ChargedTracks, GoodTracksVeryLoose, GoodTracksLoose, GoodTracksTight. The plots are made from a typical run (number 12917), normalized to display tracks/event/bin. The tagbit (see Section 5.2.8) used for generating these plots is called isPhysics, this is a very loose tagbit which essentially only requires an event to have more than two charged tracks to remove $e^{+} e^{-}$and $\mu^{+} \mu^{-}$events [26].

### 5.2.6 Composition Tools

Two packages named CompositionTools and CompositionSequences form a tool set that create all the lists of composite particles required for this analysis. CompositionTools is a set of generic particle finders which accept input lists of charged tracks, neutral tracks and/or composites and then, along with a set of loose kinematic cut parameters, they produce an output list of a particular composite particle. For example, the $D^{*}$ finder requires a list of $D^{0}$ candidates plus a list of charged tracks. The output of this finder is a list of $D^{*}$ candidates made from the input candidates. CompositionSequences provides a predefined set of input lists and parameters for CompositionTools that most members of the BaBar Collaboration find useful. This allows for a certain amount of uniformity and reduced redundancy between the efforts of various analysis groups.

The following composites and decay channels are used for this analysis [1] (the $D^{0}$ decay modes to $K \pi \pi \pi$ and $K \pi \pi^{0}$ are not currently used, but are given for reference purposes):

| Decay Mode | Particle Mass | Branching <br> fraction |
| :--- | ---: | ---: |
| $\pi^{0} \rightarrow \gamma \gamma$ | $134.977 \mathrm{MeV} / c^{2}$ | $98.798 \%$ |
| $K_{S} \rightarrow \pi^{+} \pi^{-}$ | $497.672 \mathrm{MeV} / c^{2}$ | $68.61 \%$ |
| $D^{0} \rightarrow K \pi$ | $1864.5 \mathrm{MeV} / c^{2}$ | $3.83 \%$ |
| $D^{0} \rightarrow K \pi \pi \pi$ | $1864.5 \mathrm{MeV} / c^{2}$ | $7.49 \%$ |
| $D^{0} \rightarrow K \pi \pi^{0}$ | $1864.5 \mathrm{MeV} / c^{2}$ | $13.9 \%$ |
| $D^{*} \rightarrow D^{0} \pi$ | $2010.0 \mathrm{MeV} / c^{2}$ | $67.7 \%$ |
| $B^{0} \rightarrow D^{*} \ell \nu$ | $5279.4 \mathrm{MeV} / c^{2}$ | $4.6 \%$ |

Table 5.1: Composite particles used in this analysis, provided by CompositionTools.

The $\pi^{0}$ candidate list is created by combining two particles from the GoodPhotonLoose list which satisfy the following requirements:

1. the sum of the two photon energies must be larger than 200 MeV
2. the photon pair invariant mass must be in the range of $90-170 \mathrm{MeV} / \mathrm{c}^{2}$

Photon pairs which pass these cuts form the piOLoose list. This list is refitted using the $\pi^{0}$ mass as a constraint to create the piODefaultMass list. The $\pi^{0}$ mass constraint improves the energy resolution, the Default list is the one used in this analysis.

The $D^{0}$ candidates are reconstructed in four different modes and collected into a single D0Loose list. The kaon is selected from the GoodTracksLoose list. The pion candidates are taken from the GoodTracksVeryLoose or piODefaultMass list. The following cuts are applied around the nominal (PDG) $D^{0}$ mass:

- $D^{0} \rightarrow K \pi$ : a $D^{0}$ mass window of $\pm 90 \mathrm{MeV} / c^{2}$.
- $D^{0} \rightarrow K \pi \pi \pi$ : a $D^{0}$ mass window of $\pm 90 \mathrm{MeV} / \mathrm{c}^{2}$.
- $D^{0} \rightarrow K \pi \pi^{0}$ : a $D^{0}$ mass window of $\pm 160 \mathrm{MeV} / c^{2}$.

The $D^{*}$ candidate list, DstarNeutralDLoose, is created by combining a member of the D0Loose list with a slow pion candidate from the GoodTracksVeryLoose list. The DstarNeutralDLoose candidate must meet the following requirements (the CM frame for these purposes is generally the $\Upsilon(4 \mathrm{~S})$ frame, as the $B$ has so little momentum in this frame that its effects on its daughters for loose cut purposes are small):

1. fall within a $D^{*}$ mass window of $\pm 500 \mathrm{MeV} / c^{2}$ around the nominal $D^{*}$ mass.
2. fall within a mass window of $130-170 \mathrm{MeV} / c^{2}$ for the $D^{*}-D^{0}$ mass difference.
3. have a maximum momentum of $450 \mathrm{MeV} / \mathrm{c}$ in the CM frame for the slow pion.
$D^{*}$ candidates where the $D^{0}$ decays into a charged kaon are called "Right Sign" if the slow pion and kaon have the expected charge correlation $\left(K^{+}, \pi^{-}\right.$or $\left.K^{-}, \pi^{+}\right)$. Candidates not meeting this expectation are called "Wrong Sign" candidates and are discarded from the DstarNeutralDLoose list.

The final $D^{*} \ell \nu$ candidate is created by combining the DstarNeutralDLoose list with a lepton candidate from GoodTracksLoose list. The lepton candidates are required to have a minimum momentum of $0.8 \mathrm{GeV} / \mathrm{c}$ in the CM frame, and "Wrong Sign" combinations of the lepton and slow pion are rejected.

### 5.2.7 Particle Identification

The types of particles actually directly detected and used by this analysis are $K^{ \pm}, e^{ \pm}, \pi^{ \pm}$(for the extra $D^{0}$ modes discussed in Section 5.2.2, $\gamma^{\prime}$ s from $\pi^{0}$ decays would also be used, as would $\mu^{\prime}$ 's for the $B \rightarrow D^{*} \mu \nu$ mode, of course). Various measured quantities are used to estimate the likelihood of one particular particle hypothesis or another. These likelihoods can then be used to reduce the combinatoric backgrounds generated by using incorrect particle types during reconstruction.

Particle Identification (particle ID, or just PID) uses the energy loss per distance traveled $(\mathrm{d} E / \mathrm{d} x)$ in the SVT and DCH, shower shape information from the EMC plus

Cerenkov light detected in the DIRC to formulate the particle ID hypothesis. The likelihood for each type of particle species is determined and made availible in the form of a bitmap which represents a hierarchy of probabilities. PID is most important for particles which are less likely (i.e. not pions) since these particles impact the combinatoric background most heavily. This analysis makes use of particle ID in several different ways:

1. Kaon particle ID is used to reduce combinatorics while reconstructing $D^{0 \prime}$ s.
2. Lepton particle ID is used to reduce background events while reconstructing $D^{*} \ell \nu$.
3. In Chapter 8, particle ID correction tables are used to estimate the systematic error due to incorrect determination of the particle type of a given track.

### 5.2.7.1 Kaon Particle ID

Kaon particle ID is based on information from the DCH, SVT, and DIRC. The $\mathrm{d} E / \mathrm{d} x$ information provided by the SVT yields better than a $2 \sigma$ separation between kaons and pions up to a particle momentum of about $0.6 \mathrm{GeV} / c$, and up to $0.7 \mathrm{GeV} / c$ for the DCH. For particle momentum above $1.5 \mathrm{GeV} / c$ the DCH once again gives a better than $2 \sigma$ separation due to relativistic rise [30] (see Figure 4.7 for a plot of $\operatorname{DCH} \mathrm{d} E / \mathrm{d} x$ ).

The DIRC provides a measurement of the Cerenkov angle and the number of photons arriving for each charged particle passing through the quartz bars of the DIRC. For particles with momentum $>\operatorname{mass} / \sqrt{n^{2}-1}$, where $n$ is the index of refraction for quartz, Cerenkov light will be emitted. The number of photons produced for a fixed particle path follow Poissonian statistics with a central value which depends on the particle
type, charge, momentum, polar angle and bar number. The particle momentum, individual photon Cerenkov angle, number of photons and photon arrival time are all combined in a simultaneous fit to calculate the Cerenkov angle of the charged track and the kaon likelihood [30].

Figure 5.2 shows the distribution of Cerenkov angle vs momentum and the kaon/pion separation achieved by the DIRC. The figure on the right hand side is based on the resolution of pions from a $D^{*}$ control sample. The $2.5 \sigma$ separation is slightly optimistic, though if both kaon and pion resolutions are measured the separation is approximately $2 \sigma$ at $4 \mathrm{GeV} / c$ momentum. [30]

Kaon particle ID is available to all analyses in the form of a bitmap with the following levels:

- Very Tight: designed to keep misidentification rates below $2 \%$ up to $4 \mathrm{GeV} / c$.
- Tight: designed to keep misidentification rates below $5 \%$ up to $4 \mathrm{GeV} / c$.
- Loose: designed to keep misidentification rates below $7 \%$ up to $4 \mathrm{GeV} /$ c.
- Very Loose: designed to to be highly effcient for kaons.
- Not a Pion: designed to maximize kaon efficiency while rejecting pions.


### 5.2.7.2 Electron Particle ID

Electron particle ID makes use of information from the DCH, DIRC and EMC. The DCH once again provides $\mathrm{d} E / \mathrm{d} x$ vs. momentum information. The $\mathrm{d} E / \mathrm{d} x$ for a elec-
trons is peaked (in arbitrary units) at $\approx 650$ with a width $\sigma \approx 50$. The DIRC provides information which is used only by the very tight selector. The number of measured photons should be greater than 10 and the measured Cerenkov angle is required to be within $3 \sigma$ of the electron mass hypothesis. The EMC provides five measured quantities used in electron particle ID:

- Lateral energy distribution (LAT): This is one of two shower shape variables, defined as so:

$$
\begin{equation*}
L A T=\frac{\sum_{i=3}^{n} E_{i} r_{i}^{2}}{\sum_{i=3}^{n} E_{i} r_{i}^{2}+E_{1} r_{0}^{2}+E_{2} r_{0}^{2}}, \quad E_{1} \geq E_{2} \geq \ldots \geq E_{n} \tag{5.2}
\end{equation*}
$$

where $r_{0}=5 \mathrm{~cm}$, the average distance between EMC crystal fronts, $r_{i}$ is the distance between the $i^{t} h$ crystal and the shower center, $E_{i}$ is the energy deposited in the $i^{t h}$ crystal and the sum is over all crystals in the shower.

LAT is essentially a normalized weighted sum of all energy deposited in EMC crystals by a shower excluding the two largest crystal deposits. Electromagnetic showers (those from electrons or photons) deposit most of their energy in one or two crystals, with a small amount of leakage into adjacent crystals, so LAT is expected to be smaller for electromagnetic showers than hadronic showers [29].

- Zernike moments ( $A_{n m}$ ): This is the other shower shape variable used, which exploits the fact that hadronic showers tend to be more irregular than electromagnetic showers. An expansion in angular moments can exploit this fact, and these
moments are defined as follows:

$$
\begin{align*}
& A_{n m}=\sum_{r_{i} \leq R_{0}}^{n} \frac{E_{i}}{E} \cdot f_{n m}\left(\frac{r_{i}}{R_{0}}\right) \cdot e^{-i m \phi_{i}}, \quad R_{0}=15 \mathrm{~cm}  \tag{5.3}\\
& f_{n m}(\rho)=\sum_{s=0}^{(n-m) / 2} \frac{(-1)^{s}(n-s)!\rho_{i}^{n-2 s}}{s!((n+m) / 2-s)!((n-m) / 2-s)!} \tag{5.4}
\end{align*}
$$

where $\mathrm{n}, \mathrm{m} \geq 0$, $\mathrm{n}-\mathrm{m}$ even and $\mathrm{m} \leq \mathrm{n}$. Again $E_{i}$ is the energy deposited in the $i^{t} h$ crystal and $r_{i}$ is the distance from that crystal to the shower center. A number of Zernike moments have been investigated, but the most useful one and only one currently used is the $A_{42}$ moment [29].

- E/p: The measurement of energy deposited in the calorimeter divided by the associated track momentum is an excellent means of identifying electrons. When an electron enters the calorimeter, it will produce an electromagnetic shower consisting of photon and $e^{+}, e^{-}$pairs, which deposit the energy of the original electron in the calorimeter. An ideal calorimeter will have $E / p=1$ for an electron. Various resolution effects and shower leakage will smear this distribution in a real calorimeter. The value of $\mathrm{E} / \mathrm{p}$ could even be quite a bit higher than unity if bremsstrahlung photons enter the shower, since the measured track momentum will be smaller due to the bremsstrahlung, but the deposited energy will remain the same (the EMC crystals absorb nearly all energy from electrons and photons equivalently).

Muons will deposit energy in the calorimeter as a single ionizing particle and generally exit the calorimeter only depositing a fraction of their total energy in it. Hadrons sometimes interact in the calorimeter, and sometimes pass through like
a muon, but they rarely deposit all their energy as an electron or photon does [10].

- Track-Bump separation: The path of an associated charged track is extrapolated to the front of the calorimeter and the angular position $\left(\phi_{E M C}\right)$ is recorded. The angular position of the shower cluster ( $\phi_{\text {cluster }}$ ) is also recorded. Since electromagnetic showers begin quickly as soon as particles enter the crystals, these two positions should be very close for these types of particles. Hadronic showers tend to develop later and thus the track-bump separation tends to be larger. Thus,

$$
\begin{equation*}
\Delta \phi=q \cdot\left(\phi_{E M C}-\phi_{\text {cluster }}\right), \quad \mathrm{q}=\text { charge of particle } \tag{5.5}
\end{equation*}
$$

can be used to discriminate between electrons and hadrons.

- Number of Crystals: The total number of crystals hit during a shower should be larger than two to guard against spurious background noise.

The various levels of electron particle ID are summarized in Table 5.2.

| Category | $\mathrm{d} E / \mathrm{d} x$ | $N_{\text {crystal }}$ | $\mathrm{E} / \mathrm{p}$ | LAT | $A_{42}$ | $\Delta \phi$ | DIRC |
| :--- | :--- | :--- | :---: | :---: | :---: | :---: | :---: |
| noCal | $540 \ldots 860$ |  |  |  |  |  |  |
| veryLoose | $500 \ldots 1000$ | 3 | $0.50 \ldots 5.0$ | $-10 \ldots 10$ | $-10 \ldots 10$ | no | no |
| loose | $500 \ldots 1000$ | 3 | $0.65 \ldots 5.0$ | $-10 \ldots 10$ | $-10 \ldots 10$ | no | no |
| tight | $500 \ldots 1000$ | 3 | $0.75 \ldots 1.3$ | $0 \ldots 0.6$ | $-10 \ldots 10$ | no | no |
| veryTight | $540 \ldots 860$ | 3 | $0.89 \ldots 1.2$ | $0 \ldots 0.6$ | $-10 \ldots 0.11$ | yes | yes |

Table 5.2: Electron particle ID category requirements (the $\Delta \phi$ and DIRC columns indicate whether these variables are used for the given category).

### 5.2.7.3 Muon Particle ID

Muon particle ID primarily makes use of information from the IFR. The following list describes measured quantities provided by the IFR:

- $N_{L}$ : The number of IFR layers hit in a cluster.
- $\lambda$ : The number of interaction lengths traversed by the track. This is estimated using the track extrapolated into the IFR to the last layer hit.
- $\Delta \lambda$ : The number of interaction lengths expected to be traversed by a muon is calculated, then the quantity $\Delta \lambda=\lambda_{\text {expected }}-\lambda$ is calculated.
- $T_{c}$ : This represents the continuity of a track in a cluster. It is defined as $T_{c}=$ $N_{L} /$ (Last Layer number - First Layer number). A perfectly continuous track will have $T_{c}=1$ while a track with only sporadic hits will have $T_{c}<1$.
- $\bar{m}$ : The average multiplicity of hit strips per layer.
- $\sigma_{m}$ : The standard deviation of $\bar{m}$.
- $\chi_{t r k}^{2}$ : The $\chi^{2}$ /d.o.f. of the IFR hit strips with respect to the track extrapolation.
- $\chi_{f i t}^{2}$ : The $\chi^{2}$ /d.o.f. of the IFR hit strips with respect to a third order polynomial fit of the cluster.

The amount of energy deposited in the EMC ( $E_{\text {cal }}$ ) is recorded if available. Only the Minimum Ionizing category requires there to be a value for $E_{\text {cal }}$, all other categories
drop the $E_{\text {cal }}$ cut if there is no EMC information. Muon particle ID levels are summarized in Table 5.3.

| Category | $E_{\text {cal }}$ | $N_{L}>$ | $\Delta \lambda<$ | $\lambda>$ | $T_{c}>$ | $\bar{m}<$ | $\sigma_{m}<$ | $\chi_{\text {trk }}^{2}<$ | $\chi_{\text {fit }}^{2}<$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Min. Ionizing | $0 \ldots 0.5$ | - | - | - | - | - | - | - | - |
| VeryLoose | $0 \ldots 0.5$ | 1 | 2.5 | 2.0 | 0.1 | 10 | 6 | - | - |
| Loose | $0 \ldots 0.5$ | 1 | 2.0 | 2.0 | 0.2 | 10 | 6 | 7 | 4 |
| Tight | $0.05 \ldots 0.4$ | 1 | 2.0 | 2.2 | 0.3 | 8 | 4 | 5 | 3 |
| VeryTight | $0.05 \ldots 0.4$ | 1 | 2.0 | 2.2 | 0.34 | 8 | 4 | 5 | 3 |

Table 5.3: Muon particle ID category requirements.



Figure 5.2: Left, the distribution of Cerenkov angle versus momentum for different particle types (the top band is from pions, the next one is from kaons). Right, the kaon/pion separation in units of $\sigma$.

### 5.2.8 Tagbit selection

Tagbit selection is an important process used to limit the total number of events an analysis needs to look at from the BaBar dataset. The tagbit essentially acts as a flag which indicates that a particular event has already passed a certain number of cuts. If a particular analysis uses cuts which are all as tight or tighter than a given tagbit, the analysis can save a considerable amount of computer processing time by choosing to only look at events which have that tagbit set.

All events which pass the L3 trigger are processed offline to reconstruct in detail all the tracks and clusters present in each accepted event. At this time Composition Tools is run and various particle candidates of varying quality are created. Tagbit code uses these lists and tightens cuts made in order to reduce the number of combinatoric candidates while retaining good candidates. If at least one acceptable candidate remains in the event after this tightening process, a tagbit is set 'on' for this event. In the case of $D^{*} \ell \nu$, there are five tagbits of interest to us generated during the offline processing:

- B0ToDstarlnuLoose
- B0ToDstarlnuTight
- B0ToDstarlnuVTightElec - "very tight electron sample"
- B0ToDstarlnuVTightMuon - "very tight muon sample"

The very tight electron and muon tagbits represent our signal sample. The first two tagbits are looser versions of our signal and control sample tagbits, and they are
maintained in case disaster strikes and it is discovered that the last two tagbits are somehow too aggressive. The tight or loose tagbit could be used instead of the very tight tagbits if such a problem arose. The following lists show the additional constraints that each tagbit imposes.

## - B0ToDstarlnuLoose

This combines $D^{*} \ell \nu$ candidates from the DstarAllLoosePID list and lepton candidates (with no lepton identification applied) with $p_{\ell}^{*}>0.8 \mathrm{GeV} / c$ from the GoodTracksTight list. The following $D^{0}$ decay modes are used: $K^{-} \pi^{+}, K^{-} \pi^{+} \pi^{+} \pi^{-}$, $K \pi^{+} \pi^{0}$ and $K_{S} \pi^{+} \pi^{-}$. The selection criteria are:

- $\pi^{o}$ is fit with mass constraint.
- Candidates with $K^{-} \pi^{+}, K^{-} \pi^{+} \pi^{+} \pi^{-}$and $K \pi^{+} \pi^{0}$ modes are taken from the DstarChrgKLoosePID list with:
* $D^{0}$ from the D0ChrgKLoosePID list within $\pm 90 \mathrm{MeV} / c^{2}$ mass window,
* right and wrong sign combinations.

The charged kaon is required to pass the SMSnotAPion selector.

- Slow $\pi$ from the GoodTracksVeryLoose list.
$-p_{\pi_{\text {slow }}}<450 \mathrm{MeV} / c$.
- Raw $D^{*}$ mass within $500 \mathrm{MeV} / c^{2}$ of nominal.
- $130 \mathrm{MeV} / c^{2}<m\left(D^{*}\right)-m\left(D^{0}\right)<170 \mathrm{MeV} / c^{2}$, with mass-constraint $D^{0}$.
- B0ToDstarlnuTight

A refinement of the BOToDstarlnuLoose list with additional cuts:

- $p_{l}^{*}>1.0 \mathrm{GeV} / c$,
- $D^{0}$ mass window narrowed to $\pm 40 \mathrm{MeV} / c^{2}$ for $K^{-} \pi^{+}, K^{-} \pi^{+} \pi^{+} \pi^{-}$and $K_{S} \pi^{+} \pi^{-}$ modes, and $\pm 70 \mathrm{MeV} / c^{2}$ for $K \pi^{+} \pi^{0}$ mode,
- $p_{\pi_{\text {slow }}}^{t}>50 \mathrm{MeV} / c$,
- $0.5 \mathrm{GeV} / c<p^{*}\left(D^{*}\right)<2.5 \mathrm{GeV} / c$.
- B0ToDstarlnuVTightElec

A refinement of the BOToDstarlnuTight list with additional cuts:

- $p_{l}^{*}>1.2 \mathrm{GeV} / c$,
- $D^{0}$ mass window narrowed to $\pm 20 \mathrm{MeV} / c^{2}$ for $K^{-} \pi^{+}, K^{-} \pi^{+} \pi^{+} \pi^{-}$and $K_{S} \pi^{+} \pi^{-}$ modes, and to $\pm 35 \mathrm{MeV} / c^{2}$ for $K \pi^{+} \pi^{0}$ mode,
- lepton candidate must pass very tight electron ID.


## - B0ToDstarlnuVTightMuon

Same as B0ToDstarlnuVTightElec except that lepton candidate must pass very tight muon ID.

### 5.2.9 Analysis Cuts

In this section we describe all the variables and cuts used in selecting $B^{0} \rightarrow D^{*} \ell \nu$ events. The presentation of the cuts is organized to follow the logical progression of first
creating the $D^{0}$ composite, then the $D^{*}$ and finally the $D^{*} l$ candidate. In many cases, a cut on a certain quantity is made several times at different points (during the tagbit creation, ntuple production, ascii file production, final cut selection), generally in increasing level of tightness. The plots in this section show the effect of the final cuts on a sample.

In order to show the full range of cuts performed on the data sample, two different data sets were created:

- No-Tagbit data: A sample roughly equivalent to $1 \mathrm{fb}^{-1}$ of generic data $\left(B^{0} \bar{B}^{0}+\right.$ $B^{+} B^{-}+c \bar{c}$ Monte Carlo events) was produced by which contains only the cuts imposed by Composition Tools and a lepton momentum cut of $0.8 \mathrm{GeV} / c$. This represents essentially the loosest cuts that can be applied. The momentum cut could be relaxed further but this adds very few signal events while introducing an extremely large number of combinatorics to the $D^{*}$-lepton list. The purpose of this sample is to study the effects of cuts applied to $D^{0}$ and $D^{*}$ candidates, which cannot be done when events are selected with tight tagbits.
- Control Sample: This is an approximately $9 \mathrm{fb}^{-1}$ sample of unfiltered $B^{0} \bar{B}^{0}, B^{+} B^{-}$ and $c \bar{c}$ events made in the proper ratios (as in Section 5.2.3). All $D^{0}$ and $D^{*}$ cuts plus a $1.2 \mathrm{GeV} / \mathrm{c}$ lepton momentum cut have been applied to all events. This sample is useful for studying the cuts used to further refine the $D^{*} \ell \nu$ candidate list.

The great utility of Monte Carlo data is that the underlying decay is known exactly. A one-to-one correspondence can generally be created between detected tracks and Monte Carlo truth tracks. When a detected track has a corresponding Monte Carlo truth
track, the tracks are called "MC-matched". The Monte Carlo truth matching algorithm used by BaBar is extremely efficient for correctly matching charged tracks with a Monte Carlo truth partner, though the same is not true for neutral tracks.

Neutral tracks are much harder to match since there is considerably less information gathered on neutral tracks. The primary detector for neutrals is the EMC while charged tracks can leave information in the SVT, DCH, DIRC, EMC and even the IFR for muons and the neutral $K_{L}$ particles. This weakness fortunately has very little effect on this analysis. The only possible problem occurs in the $D^{0} \rightarrow K \pi \pi^{0}$ mode, but there exist enough information in the parent $D^{0}$ and charged daughters to overcome any possible problems in MC matching. The neutrino is of course never MC-matched since it is never detected.

### 5.2.9.1 $\quad D^{0}$ Cuts

The first composite particle created is the $D^{0}$. The summary of the cuts on the $D^{0}$ described below is shown in Table 5.4.

| $D^{0}$ cut | $K \pi$ | $K \pi \pi \pi$ | $K \pi \pi^{0}$ | $K_{S} \pi \pi$ |
| :--- | :---: | :---: | :---: | :---: |
| Mass cut $\left(1864.5 \mathrm{MeV} / c^{2}\right)$ | $\pm 17 \mathrm{Mev}$ | $\pm 17 \mathrm{MeV} / c^{2}$ | $\pm 34 \mathrm{MeV} / c^{2}$ | $\pm 17 \mathrm{MeV} / c^{2}$ |
| $D^{0}$ vertex | 0.001 | 0.001 | 0.001 | 0.001 |
| Kaon particle ID | Not A Pion | Tight | Tight | none |
| Dalitz cut |  |  | 0.1 | 0.1 |
| $K_{S}, \pi^{0}$ vertex |  |  | 0.01 | 0.01 |
| $K_{S}, \pi^{0}$ mass cut |  |  | $134.977 \pm 15.75 \mathrm{MeV} / c^{2}$ | $497.672 \pm 15 \mathrm{MeV} / c^{2}$ |

Table 5.4: Summary of $D^{0}$ cuts.

The first cut made on the $D^{0}$ candidates is a mass cut. Figure 5.3 shows the $D^{0}$ masses for all candidates in red, the blue plot indicates $D^{0 \prime}$ s which are matched to a real Monte Carlo $D^{0}$, and the black dotted lines show the value of the final $D^{0}$ mass cut value, a list of all cut values is complied in the final cut section.

The quality of the vertex fit is also used to remove background events. The $\chi^{2}$ and degrees of freedom of the fit are converted into a probability and events with a probability of the $D^{0}$ daughters vertexing to a common point being less then 0.001 are rejected, see Figure 5.4. Tight Kaon particle ID also is used to eliminate combinatoric $D^{0 \prime}$ s in the $D^{0} \rightarrow K \pi \pi \pi$ and $D^{0} \rightarrow K \pi \pi^{0}$ modes. NotAPion Kaon particle ID is used in the $D^{0} \rightarrow K \pi$ mode, since it is already very clean. Kaon particle ID is not available for the $K_{S}$ mode. Figure 5.5 shows the number of events eliminated by the Kaon particle ID cut while Figure 5.6 shows the number of MC matched events lost due to this cut.

The decay mode $D^{0} \rightarrow K \pi \pi^{0}$ includes a $\pi^{0}$ daughter and has has extra backgrounds produced by these neutral particles. The $D^{0} \rightarrow K \pi \pi^{0}$ decay has several resonance states which contribute to the decay. The most important resonance is $D^{0} \rightarrow$ $K^{-} \rho^{+}$, but the $D^{0} \rightarrow K^{*-} \pi^{+}$and $D^{0} \rightarrow K^{* 0} \pi^{0}$ are also significant. The Dalitz plot densities are calculated with the three-body decay kinematic variables, $m_{K^{-} \pi^{+}}^{2}, m_{K^{-} \pi^{0}}^{2}$, $m_{K^{-} \pi^{+} \pi^{0}}^{2}$. The Dalitz plot distribution at generator level and contours is shown in Figure 5.7. This distribution is then convoluted with $m_{K^{-} \pi^{+}}^{2}$ and $m_{K^{-} \pi^{0}}^{2}$ resolution functions to produce the final density plot. The plot is normalized such that the largest value possible is unity. Events with a value outside the 0.1 contour are rejected.

Figure 5.8 shows the mass distributions for the $\pi^{0}$ (and $K_{S}$, for the no longer used $D^{0} \rightarrow K_{S} \pi \pi$ mode).

Finally, the $\pi^{0}$ also has a pseudo-vertex from its photon daughters, see Figure 5.9 (not a true vertex, this is actually a kinematic fit, as the photons leave no tracks in the SVT or DCH, but their trajectories are traced back to near the interaction point from their directions as obtained by their cluster hits in the EMC).

In sum, these cuts significantly reduce the number of combinatoric $D^{0 \prime}$ s while retaining the majority of good candidates.


Figure 5.3: A plot of $D^{0}$ masses for all $D^{0}$ candidates (red dashed) and MC matched candidates (solid blue) for the four $D^{0}$ decay modes. The final cut values are shown by the black vertical lines.


Figure 5.4: Top: Plots of $D^{0}$ vertex probabilities for all $D^{0}$ candidates (left) and MC matched candidates (right). Bottom left: This plot shows the effect of the D0Vtx cut on all events - the red dashed curve is all events while the solid blue curve is events passing the $D^{0}$ vertex cut. Bottom right: This shows the effect of the cut on MC-matched events. The red dashed curve is all MC-matched events before the cut, the solid blue curve is MC-matched events after the cut.


Figure 5.5: The effect of applying Kaon ID to the the $D^{0}$ samples. The red dashed curve shows all events, the solid blue curve is all events which pass Kaon ID. The $K \pi$ mode uses the "not a pion" requirement which has virtually no effect at all, the $K^{-} \pi^{+} \pi^{+} \pi^{-}$ (left) and $K \pi^{+} \pi^{0}$ (right) modes use the "tight kaon" particle ID, these two modes are shown above.


Figure 5.6: The effect of applying Kaon ID to the the $D^{0}$ samples on MC-matched events. The red dashed curve shows all MC-matched events, the solid blue curve shows the MCmatched events remaining after the Kaon ID cut ( $K^{-} \pi^{+} \pi^{+} \pi^{-}$mode (left) and $K \pi^{+} \pi^{0}$ mode (right)).


Figure 5.7: Top: $D^{0} \rightarrow K \pi \pi^{0}$ Dalitz distribution at the generator level without convolving the distribution with resolution functions. Bottom: The region which falls inside the typical cut based region of $\left|m\left(\pi^{+} \pi^{-}\right)-m(\rho)\right|<250 \mathrm{MeV} / c^{2}$ and $\left|\cos \theta_{K \pi^{+}}^{*}\right|>0.4$ in the $\pi^{+} \pi^{-}$rest frame.


Figure 5.8: Left: $K_{S}$ mass plot (for the currently unused $D^{0} \rightarrow K_{S} \pi \pi$ mode), the dashed red curve is all events, solid blue curve is MC-matched events. Right: $\pi^{0}$ mass plot (for the $D^{0} \rightarrow K \pi \pi^{0}$ mode), the dashed red curve is all events, the solid blue curve is MCmatched events.




PiOVtx Prob MC matched Pi0s



Figure 5.9: Top: The left plot shows all $\pi^{0}$ events, the right plot shows all MC-matched $\pi^{0}$ events. The vertex cut applied on $\pi^{0}$ is $\operatorname{prob}\left(\chi^{2}\right)>0.01$ (the first bin on both plots). Bottom left: shows all events (red dashed) and events passing the vertex cut (solid blue). Bottom right: the red dashed curve is all MC-matched $\pi^{0}$ events, the solid blue curve is all MC-matched events which pass the vertex cut.

### 5.2.9.2 $D^{*}$ Cuts

This stage of reconstruction involves adding a single charged particle from the GoodTracksVeryLoose list to the $D^{0}$ candidate. Since the $D^{0}$ mass plus $\pi$ mass almost add up to the $D^{*}$ mass ( $D^{*}$ mass - $D^{0}$ mass $-\pi$ mass $\approx 5 \mathrm{MeV} / c^{2}$ ), the $D^{0}$ and $\pi$ are produced nearly at rest in the $D^{*}$ frame. The $D^{0}$ and $\pi$ have almost the exact same velocity as the parent $D^{*}$ which means that the $D^{0}$ and $\pi$ split the parent momentum in the same ratio as the ratio of their masses $(1865 / 139 \approx 13)$. Thus the $\pi$ will have a small momentum in the lab frame, which is where the label "slow pion" comes from.

There are two quality cuts made to improve the purity of the $D^{*}$ sample. The first is a cut of $50 \mathrm{MeV} / c$ on the minimum transverse momentum of the slow pion, as shown in Figure 5.10. The second is a cut on the $D^{*}$ momentum in the center of mass frame, shown in Figure 5.11. The lower cut of $0.5 \mathrm{GeV} / \mathrm{c}$ is close to the kinematic lower limit for good $D^{* \prime}$ s, the $2.5 \mathrm{GeV} / c$ upper limit cuts out events which have a good $D^{*}$ but those not coming from $B^{0} \rightarrow D^{*} \ell \nu$.


Figure 5.10: Left: the momentum distribution of the slow pion. The red dashed curve represents all $D^{*}$ candidate, the solid blue curve is MC-matched $D^{* \prime}$. Right: Same as figure on the left except the slow pion candidate is required to have at least 5 DCH hits.

## CM Dstar Momentum



Figure 5.11: The momentum of the $D^{*}$ in the center of mass reference frame. The red dotted curve represents all $D^{*}$ candidates, the blue dashed curve is all MC-matched $D^{*}$ candidates while the solid green curve shows all MC-matched $D^{*} \ell \nu$ candidates (this shows especially the effectiveness of the upper limit cut).

### 5.2.9.3 $\quad D^{*} \ell \nu$ Cuts

In addition to the standard vertex probability and momentum spectrum variables (for the $D^{*} \ell$ and lepton, respectively), there are several new variables at this stage which help improve the quality of the signal sample:

- Thrust: The angle, $\theta_{\text {thrust }}$, is the angle between the thrust axis of the $D^{*} l$ candidate and the thrust axis of the rest of the event. Continuum background events tend to be very jetty and have a $\left|\cos \theta_{\text {thrust }}\right|$ close to one.
- $\cos \theta_{B, D^{*} l}\left(\right.$ also called $\left.\cos \theta_{B Y}\right)$ : The cosine of the angle between the $B$ direction and the $D^{*}+$ lepton. This angle comes from using the 4 -vector equation $P_{B}=P_{D^{*} l}+P_{\nu}$. If the $P_{D^{*} l}$ term is subtracted from both sides then both sides squared yields:

$$
\begin{gather*}
0=m_{B}^{2}+m_{D^{*} l}^{2}-2\left(E_{B} E_{D^{*} l}-\left|p_{B}\right|\left|p_{D^{*} l}\right| \cos \theta_{B, D^{*} l}\right)  \tag{5.6}\\
\cos \theta_{B, D^{*} l}=\frac{2 E_{B} E_{D^{*} l}-m_{B}^{2}-m_{D^{*} l}^{2}}{2\left|p_{B}\right|\left|p_{D^{*} l}\right|} \tag{5.7}
\end{gather*}
$$

Since each term is either known ahead of time ( $m_{B}$ from the PDG, $E_{B}, p_{B}$ from beam information) or measured ( $m_{D^{*} l}, p_{D^{*} l}$ ) this angle should only take on physical values $\left(\left|\cos \theta_{B, D^{*} l}\right| \leq 1\right)$. Due to resolution effects the actual cut taken is $\left|\cos \theta_{B, D^{*} l}\right| \leq 1.2$.

- $\cos \theta_{D^{*} l}$ : Though this cut is not used in our analysis because it would cut out too many events close to $w=1$ (nearly stationary $D^{*}$ ) which are important for slope extrapolation in form factor and $\left|V_{c b}\right|$ measurements, we still describe it here because it is a commonly used cut variable in other $D^{*} \ell \nu$ analyses.

The angle between the $D^{*}$ and lepton candidate is termed $\cos \theta_{D^{*} l}$. Because the $B^{0}$ is a spin zero particle the total spin of the $D^{*}+\ell+\nu$ system must be zero. Because the $D^{*}$ is spin 1, the lepton-neutrino system will try to align its spin angular momentum to cancel the total angular momentum projected on the decay axis. With the helicity of a neutrino being fixed (i.e., left-handed neutrinos, right handed anti-neutrinos), the lepton tends to come out back-to-back from the $D^{*}$ with a harder spectrum than the neutrino. A number of backgrounds will have the same topology: $c c b a r$ events will have a $D^{*}$ and lepton essentially back to back as will fake lepton events ( $B^{0}$ two body decay). However, the one type of background this cut would be effective against if used is the uncorrelated lepton (from the opposite-side $B$ ), which has a flat distribution in $\cos \theta_{D^{*} l}$ before the $\cos \theta_{B, D^{*} l}$ cut is made.

To summarize, good $D^{*}$ and $D^{*} \ell \nu$ candidates are required to pass the cuts shown in Table 5.5.

| $D^{*}$ cut | Cut value |
| :--- | :---: |
| Slow pion momentum | $0.05<\mathrm{p}<0.45 \mathrm{GeV} / c$ |
| Momentum of $D^{*}$ | $0.5<p_{D^{*}}^{*}<2.5 \mathrm{GeV} / c$ |
| $D^{*} \ell \nu$ cut |  |
| $D^{*} \ell \nu$ vertex probability | 0.01 |
| $P_{\text {lepton }}^{*}$ | $1.2 \mathrm{GeV} / c$ |
| Lepton particle ID | Very Tight electron or muon particle ID |
| Thrust cut | $\left\|\cos \theta_{\text {thrust }}\right\| \leq 0.85$ |
| $\cos \theta_{B, D^{*} l}$ | $-1.2 \leq \cos \theta_{B, D^{*} l} \leq 1.2$ |

Table 5.5: Summary of $D^{*}$ and $D^{*} \ell \nu$ cuts.

The first three plots of this subsection ( $D^{*} \ell$ vertex, $p_{l}^{*}$ spectrum, $\ell$ ID) use the special $1 \mathrm{fb}^{-1}$ sample (see Section 5.2.9 for the description of this sample) with all $D^{0}$ and $D^{*}$ cuts applied, the remaining plots show variables which are not affected by DstarlnuUser ntuple level cuts, which means that the $\sim 9 \mathrm{fb}^{-1}$ control sample can be used to show the effect of the remaining cuts.

Figure 5.12 shows the vertex cut applied to the $D^{*} \ell \nu$ vertex and its effect on combinatoric backgrounds.

Figure 5.13 shows the center of mass momentum of the lepton candidate for all $D^{*} \ell \nu$ candidates and for the MC-matched $D^{*} \ell \nu$ candidates. This cut is very effective at not only removing large combinatorics, but also removing cascade or secondary leptons which are real leptons with a softer spectrum than primary leptons.

Figure 5.14 shows the impact of using very tight lepton particle ID. This cut removes large amounts of background while retaining most of the MC-matched candidates.

Figure 5.15 is made from the $\approx 9 \mathrm{fb}^{-1}$ generic+ccbar sample with all DstarlnuUser cuts (everything up to this point) applied. The plots also contain a cut on the $D^{*}$ $D^{0}$ mass of $0.1454 \pm 0.0025 \mathrm{GeV} / c^{2}$, which is the signal region in $D^{*}-D^{0}$ mass space. The rest of the plots in this section (Figures 5.16 and 5.17) are made with the DstarlnuUser cuts, the $D^{*}-D^{0}$ mass window cut and a cut of $\left|\cos \theta_{\text {thrust }}\right|<0.85$.


Figure 5.12: The vertex quality cut applied to the $D^{*} l$ vertex. Top: The left plot represents all events while the right plot corresponds to MC matched events. Bottom: These are $\delta m$ plots. The left shows all events (red dashed) and events passing the cut (solid blue) while the right shows MC-matched events (red dashed) and MC-matched events passing the vertex cut (solid blue).

## Pstarlepton



Figure 5.13: The momentum of the lepton candidate in the center of mass frame. The red dashed curve shows all candidates while the solid blue curve shows MC-matched $D^{*} \ell \nu$ candidates.


Figure 5.14: Left: the red dashed curve shows all events while the solid blue curve shows all events passing the good lepton cut (i.e., passes very tight lepton ID). Right: These are $\delta m$ plots. The red dashed curve shows all MC-matched events and the solid blue curve shows all MC-matched events passing the good lepton cut.


Figure 5.15: The distribution of the thrust variable broken down by different types of backgrounds. Upper left: the black dashed line (topmost line) shows the thrust distribution after all $D^{0}$ and $D^{*}$ cuts have been made (including the $D^{*}-D^{0}$ mass cut). The solid blue line shows MC-matched $D^{*} \ell \nu$ events, the red dashed (lowest line) shows the combinatoric background (it is modest because of the mass cut). The other plots show the ccbar, flipped (direction-reversed, momentum magnitude kept constant)/ cascade (seconadary) lepton and charged $B$ plots.


Figure 5.16: A plot of $\cos \theta_{B, D^{*} l}$ for all events, MC-matched $D^{*} \ell \nu$ signal events and flipped lepton events (i.e. a real $D^{*}$ from one $B$ and a real lepton from the other), after all cuts except for this one $\left(\cos \theta_{B, D^{*} l}\right)$.


Figure 5.17: A plot of $\cos \theta_{D^{*} l}$, not cut on for this analysis.

### 5.2.10 Cut efficiency

This section provides a running count of how many events are removed, both signal and background, as a result of various cuts applied to the no-tagbit sample. Because this study was done with all four $D^{0}$ modes, we have left them in for comparison purposes.

The cuts considered are the $D^{0}$ cuts, $D^{*}$ cuts and the $D^{*} \ell \nu$ cuts. Tables 5.8 and 5.9 show the statistics for each level of applied cuts (these numbers are only meant to be schematic, as they include cuts such as $\cos \theta_{D^{*} \ell}$ which we do not make in this analysis, and which modify the final results, but only slightly).

The branching fractions (BF's) used by BaBar Monte Carlo do not always match the PDG [1] values, particularly for modes with branching fractions given only as upper limits. The Monte Carlo generator requires that all the branching fractions add to one, thus some fractions are scaled to meet this requirement. These fractions can change from software release (i.e., version) to release, Table 5.6 shows the exact numbers used for this batch of Monte Carlo data (again, the $D^{0}$ decay modes to $K \pi \pi \pi, K \pi \pi^{0}$, and $K_{S} \pi \pi$ are not used in this analysis, but are given for reference purposes, as are the $\pi^{0}$ and $K_{S} \quad \mathrm{BF}$ 's):

The numbers given in Table 5.6 indicate that $2.06 \%$ of the generic $B^{0 \prime}$ s will decay into one of the four decay chains used for the study of this section (we would find just $0.6 \%$ decaying into the $D^{0} \rightarrow K \pi$ mode of the analysis). The no-tagbit sample has exactly 512,000 generic $B^{0} \bar{B}^{0}$ events, which should yield $21,075 \pm 145 \mathrm{MC}$-matched events for all four $D^{0}$ modes (an extra factor of 2 is included since either $B^{0}$ could decay as $D^{*} \ell \nu$ ). The

| Decay Mode | Branching fraction |
| :--- | :---: |
| $\pi^{0} \rightarrow \gamma \gamma$ | $98.80 \%$ |
| $K_{S} \rightarrow \pi^{+} \pi^{-}$ | $68.61 \%$ |
| $D^{0} \rightarrow K \pi$ | $3.83 \%$ |
| $D^{0} \rightarrow K \pi \pi \pi$ | $7.49 \%$ |
| $D^{0} \rightarrow K \pi \pi^{0}$ | $13.9 \%$ |
| $D^{0} \rightarrow K_{S} \pi \pi$ | $2.7 \%$ |
| $D^{*} \rightarrow D^{0} \pi$ | $68.3 \%$ |
| $B^{0} \rightarrow D^{*} \ell \nu$ | $5.6 \%$ |

Table 5.6: Branching fractions used by EvtGen (release 8.8.0c-physics-1).
actual number found in the no-tagbit sample is 4719 ( $22.4 \%$ of total yield). The difference is due to events falling outside detector acceptance, track inefficiency (especially the slow pion) and the unavoidable lepton momentum cut at $0.8 \mathrm{GeV} / c$. The efficiencies are calculated in Tables 5.7, 5.8 and 5.9.

| Decay <br> Mode | MC truth <br> expected | Actual <br> events | acceptance <br> fraction |
| :--- | :---: | :---: | :---: |
| $K \pi$ | 3000 | 1177 | $39.2 \%$ |
| $K \pi \pi \pi$ | 5867 | 1378 | $23.5 \%$ |
| $K \pi \pi^{0}$ | 10757 | 1827 | $17.0 \%$ |
| $K_{S} \pi \pi$ | 1451 | 337 | $23.2 \%$ |
| Total | 21075 | 4719 | $22.4 \%$ |

Table 5.7: Calculating the product of detector acceptance, track efficiency and Composition Tools cuts efficiencies, by decay mode.

Figures 5.18, 5.19, and 5.20 show the amount of signal and background events left after various cuts. Figure 5.21 shows the total number of MC-matched $D^{*} \ell \nu$ events prior to any cuts and the amount remaining after all cuts have been applied. All the plots in this section have been made with the special $1 \mathrm{fb}^{-1}$ sample (see Section 5.2.9).

|  | No Cuts |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Decay <br> Mode | signal <br> events | back <br> events | sig eff | S/B | signal <br> events | back <br> events | sig eff | S/B |  |  |  |  |  |  |
| $K \pi$ | 1177 | 26,653 | $39.2 \%$ | 0.044 | 1050 | 14,587 | $35.0 \%$ | 0.072 |  |  |  |  |  |  |
| $K \pi \pi \pi$ | 1378 | 406,127 | $23.5 \%$ | 0.003 | 1010 | 58,176 | $17.2 \%$ | 0.017 |  |  |  |  |  |  |
| $K \pi \pi^{0}$ | 1827 | 604,501 | $17.0 \%$ | 0.003 | 823 | 30,137 | $7.7 \%$ | 0.027 |  |  |  |  |  |  |
| $K_{S} \pi \pi$ | 337 | 317,065 | $23.2 \%$ | 0.001 | 61 | 4,455 | $4.2 \%$ | 0.014 |  |  |  |  |  |  |
| Total | 4719 | $1,354,346$ | $22.4 \%$ | 0.003 | 2944 | 106,355 | $14.0 \%$ | 0.028 |  |  |  |  |  |  |
|  | $D^{*}$ Cuts |  |  |  |  |  |  |  |  |  | $D^{*} \ell$ Cuts |  |  |  |
| $K \pi$ | 1041 | 10,738 | $34.7 \%$ | 0.097 | 373 | 170 | $12.4 \%$ | 2.194 |  |  |  |  |  |  |
| $K \pi \pi \pi$ | 1004 | 48,665 | $17.1 \%$ | 0.021 | 383 | 842 | $6.5 \%$ | 0.455 |  |  |  |  |  |  |
| $K \pi \pi^{0}$ | 813 | 24,056 | $7.6 \%$ | 0.034 | 314 | 492 | $2.9 \%$ | 0.638 |  |  |  |  |  |  |
| $K_{S} \pi \pi$ | 59 | 3,815 | $4.1 \%$ | 0.015 | 28 | 43 | $1.9 \%$ | 0.651 |  |  |  |  |  |  |
| Total | 2917 | 87,274 | $13.8 \%$ | 0.033 | 1098 | 1,547 | $5.2 \%$ | 0.229 |  |  |  |  |  |  |

Table 5.8: Summary of no cut, $D^{0}$ cut, $D^{*}$ cut and $D^{*} \ell \nu$ cut efficiency. This table list the number of signal and background events remaining after the associated cuts have been applied. Note that there is no $D^{*}-D^{0}$ cut applied (see next table for that cut).

|  | $D^{*} \ell \nu$ Cuts $+D^{*}-D^{0}$ mass cut $\left( \pm 2.5 \mathrm{MeV} / c^{2}\right)$ |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Decay | MC truth | Within | signal | bkgd | sig eff | S/B |
| Mode | expected | acceptance | events | events |  |  |
| $K \pi$ | 3000 | 1177 | 354 | 67 | $11.8 \%$ | 5.28 |
| $K \pi \pi \pi$ | 5867 | 1378 | 363 | 190 | $6.2 \%$ | 1.91 |
| $K \pi \pi^{0}$ | 10757 | 1827 | 306 | 199 | $2.8 \%$ | 1.54 |
| $K_{S} \pi \pi$ | 1451 | 337 | 27 | 9 | $1.9 \%$ | 3.00 |
| Total | 21075 | 4719 | 1050 | 465 | $5.0 \%$ | 2.258 |

Table 5.9: Summary of no cut, $D^{0}$ cut, $D^{*}$ cut and $D^{*} \ell \nu$ cut efficiency. This table list the number of signal and background events remaining after the associated cuts have been applied. Note that there is no $D^{*}-D^{0}$ cut applied (see next table for that cut).





Figure 5.18: The red dashed curve shows all $D^{*} l$ candidate events with no cuts applied. The blue dotted curve shows all real $D^{* \prime}$ s in this sample while the solid black curve shows all real $D^{*} \ell \nu$ events in this sample (shown for the four $D^{0}$ decay modes).




D*-D0 mass - kspi


Figure 5.19: The red dashed curve shows all $D^{*} l$ candidate events with only $D^{0}$ cuts applied. The blue dotted curve shows all real $D^{* \prime}$ s in this sample while the solid black curve shows all real $D^{*} \ell \nu$ events in this sample (shown for the four $D^{0}$ decay modes).





Figure 5.20: The red dashed curve shows all $D^{*} \ell \nu$ candidate events with all $D^{*} \ell \nu$ cuts applied. The solid blue curve shows all real $D^{*} \ell \nu$ events in this sample (shown for the four $D^{0}$ decay modes).


Figure 5.21: Red dashed shows all MC-matched events prior to cuts, solid blue shows MC-matched events after all cuts applied. This plot shows the efficiency for finding $D^{*} \ell \nu$ events in each decay mode.

### 5.2.11 Summary of Final Cuts

To review, and summarize in one place we list here our final event selection criteria (we first list the cuts for our decay chain only, and then either parenthesize or list additionally those cuts made for the other $D^{0}$ modes that will be used in a future analysis):

Total event level:

- Thrust cut: $\left|\cos \theta_{\text {thrust }}^{*}\right|<0.85$ where $\theta_{\text {thrust }}^{*}$ is the angle between the thrust of the $D^{*} \ell$ candidate and that of the rest of the event (as one example of how the cuts suppress the background, see Figure 5.15 for effectiveness of this specific cut. See e.g. pgs. 77-105 of [5] for many more plots of a similar type.).

$$
D^{0}:
$$

- For the charged daughters of the $D^{0}$ candidates, the $\pi^{\prime}$ s are selected from GoodTracksLoose, and the $K$ is selected from GoodTracksLoose for the $K \pi$ mode (also true for the $K \pi \pi \pi, K \pi \pi^{0}$ modes).
- The charged kaon is required to pass the notAPion criterion for the $K \pi$ mode using KaonSMSSelector.
- The raw $D^{0}$ mass is required to be within $\pm 17 \mathrm{MeV} / c^{2}$ for $K \pi$ (and for the $K \pi \pi \pi$ and $\left.K_{S} \pi \pi\right)$ of the PDG $D^{0}$ mass. The $D^{0}$ is vertexed with mass contraint and the $\chi^{2}$ vertex probability is required to be greater than $0.1 \%$.

Kaon:

- The charged kaon is required to pass the Tight criterion for the $K \pi \pi \pi$ and $K \pi \pi^{0}$ modes using the KaonSMSSelector.

$$
D^{*}:
$$

- The slow $\pi$ is selected from the GoodTracksVeryLoose list.
- The slow $\pi$ is required to satisfy $p_{\pi_{\text {slow }}}<450 \mathrm{MeV} / c$, and $p_{\pi_{\text {slow }}}^{t}>50 \mathrm{MeV} / c$.
- The momentum of the $D^{*}$ in the CM frame is required to satisfy $0.5 \mathrm{GeV} / c<p_{D^{*}}^{*}<$ $2.5 \mathrm{GeV} / c$.
- The $D^{*} \ell$ vertex is fitted with beamspot contraint. The $\chi^{2}$ vertex probability is required to be greater than $1 \%$.
- The $D^{*} \ell$ candidate satisfies $\left|\cos \theta_{B-D^{*} \ell}^{*}\right|<10$.


## Electron:

- The electron candidate is selected from the GoodTracksLoose list.
- The electron momentum in the CM frame satisfies $s p>1.2 \mathrm{GeV} / c$

Final overall cuts:

- The $D^{*}-D^{0}$ mass difference $\delta m$ before $D^{*} \ell$ refitting (but using mass-constrained $D^{0}$ ) is is required to be less than $170 \mathrm{MeV} / c^{2}$.
- The refit (attempting to fit the electron, $\pi_{s}$ and $D^{*}$ to a common vertex, after initially finding the vertex via a electron- $D^{*}$ vertex fit, then fixing the vertex and trying to fit the $\pi_{s}$ to it) for electron, $\pi_{\text {slow }}$ and $D^{0}$ daughters converges.

At the very final processing stage:

- The electron candidate passes the veryTight $e$ selector. For fake samples, the $e$ candidate fails the $e$ selector.
- We select a candidate in the refitted $142.9 \mathrm{MeV} / c^{2}<\delta m<147.9 \mathrm{MeV} / c^{2}$ mass difference signal window.
- The $D^{*} \ell$ candidate satisfies $\left|\cos \theta_{B-D^{*} \ell}^{*}\right|<1.2$.

Other cuts for the $D^{0}$ decay modes to $K \pi \pi \pi, K \pi \pi^{0}$, and for muons (for reference, and to be used in the future):

- The $\pi^{0}$ is reconstructed from two photons with raw invariant mass within $\pm 15.75$ $\mathrm{MeV} / c^{2}$ of the PDG $\pi^{0}$ mass. The $\pi^{0}$ is fit with a mass contraint. $1 \%$.
- For the $K \pi \pi^{0}$ the Dalitz probability density is required to be greater than 0.1 (we calculate the decay amplitude squared based on measurements of amplitudes and phases by E687[21] and the four-momenta of the $D^{0}$ decay products in our data, assuming perfect resolutions). The maximum is normalized to unity. See BAD 34 [26], Version 8, pp. 32-34, for details.
- For $\mu$ samples, the muon candidate passes the veryTight $\mu$ selector. For fake samples, the muon fails the $\mu$ Loose selector.


## Chapter 6

## Fitting Method

Now that we have seen how we select as pure a sample as possible of candidate signal events, we turn to how we analyze this sample of events and extract the form factor parameters from it.

### 6.1 Motivation for Moments Expansion Fitting Method

In order to translate the formalism of Chapter 3 into a form where we can measure the form factors in an experimental situation, we need to account for the fact that we will never have a pure unadulterated signal distribution of events to which we can fit the theoretical PDF, but that in fact in the real world the detector does not see all particles, cuts on certain variables must be made to suppress backgrounds etc. and we must correct for all of this. Put another way, the first step in extracting form factors from a real dataset is that we must allow and correct for the acceptance, inefficiency, and resolution
of a real detector and experimental situation. Doing this "acceptance correction" (which is the overall term we lump all these effects under) is the primary concern of this section.

We shall here explain the "Moments Method" of expanding the PDF in a basis of functions of the form factors to be measured, which is an unusual approach to take care of acceptance correction, and to our knowledge this analysis is the first time a moments expansion method has been generalized from the original form as presented in [33] to a system with arbitrary momentum transfer (i.e., with one of the initial two daughter particles being virtual, and thus off-shell). This triples the number of needed moments from 6 to 18 from the original method shown in [33], as we will see later in this chapter.

The Moments Method described and validated in the rest of this chapter has one main advantage over the Reweighting Method for working with a fixed set of MC statistics. This is the following: as we will see in the next several sections, this method requires only knowledge of the integrated efficiency function, not knowledge of the detailed behavior of the specific efficiency function in the 4-dimensional kinematic variable space (in fact, this also relies on the fact that the smearing function is narrow as we will see later in Section 6.2.1, but this is the case for our observables). This amounts to needing to know less from each MC event, so that in fact, the error due to MC statistics will behave as $\frac{\lambda}{\sqrt{\rho_{\text {MC/D }}}} \sigma_{D}$, where $\lambda$ is a proportionality factor less than unity. We have not made a rigorous test of this behavior, this would most likely require doing the analysis also in the Reweighting method and comparing the MC statistical error, but we certainly believe that our MC statistical error behavior can be no worse than that in the Reweighting

Method (we describe the estimate for our MC statistical error in detail in Section 8.2.2).
To continue with the workings of the method itself, we go back to Ref. [33] which describes the analysis of the decay of $B^{0} \rightarrow J / \psi K^{*}$ - this is analogous to our initial decay of $B^{0} \rightarrow W^{*} D^{*}$ (where the $W^{*}$ is virtual) as both are decays of a pseudoscalar to two vectors - except in the former case both final state particles are on-shell so that the momentum transfer is a fixed quantity, where it is variable in our case (see Section 3.1). In the $J / \psi K^{*}$ analysis a similar method based on the moments expansion but only requiring measurement of the three analogous angular variables is used, at a fixed $w$. The inclusion of this extra degree of freedom (variable $w$ ) highly complicates the analysis, increasing the number of moments three-fold, as mentioned above.

To illustrate the method, we show how it works first for the case of a perfect detector, and later extend it to the case of a realistic one.

### 6.1.1 Break PDF into Products of Helicity Amplitudes and Angular Pieces

To begin, define $\vec{x} \equiv\left\{\cos \theta_{\ell}, \cos \theta_{V}, \chi\right\}=$ the set of the three angular kinematic variables and $\vec{\mu} \equiv\left\{R_{1}, R_{2}, \rho^{2}\right\}=$ the parameters of the underlying distribution which we are attempting to extract.

Then define for the $j^{\prime}$ th event the PDF $g_{j}$, taken from the original PDF eq.(3.5), and rewritten as the sum of the six terms which are products of the helicity amplitudes multiplied by the angular coefficent pieces:

$$
\begin{equation*}
g_{j}\left(\vec{x}_{j}, w_{j} ; \vec{\mu}\right)=\sum_{a=1}^{6} \mathcal{H}_{a}\left(w_{j} ; \vec{\mu}\right) \Xi_{a}\left(\vec{x}_{j}\right) \tag{6.1}
\end{equation*}
$$

with

$$
\begin{equation*}
\mathcal{H}_{a}\left(w_{j} ; \vec{\mu}\right)=\left\{H_{+}^{2}, H_{-}^{2}, H_{0}^{2}, H_{+} H_{-}, H_{+} H_{0}, H_{-} H_{0}\right\} \tag{6.2}
\end{equation*}
$$

where $\mathcal{H}_{a}\left(w_{j} ; \vec{\mu}\right)$ refers to one of the set of the six biproducts of the three helicity amplitudes, (which, as we have noted in Chapter 3, are each functions of $w$, though we have suppressed that argument above in eq.(6.2) ). Thus the biproducts depend only on $w$ and simple products of the individual $\vec{\mu}$ factors, as do each of the individual helicity amplitudes themselves.

The angular coefficent pieces $\Xi_{a}(\vec{x})$ are defined as the respective angular function coefficients for each helicity amplitude term, from eq.(3.5):

$$
\begin{array}{ll}
\Xi_{++}=\sin ^{2} \theta_{V}\left(1-\cos \theta_{\ell}\right)^{2}, & \Xi_{+-}=-2 \sin ^{2} \theta_{\ell} \sin ^{2} \theta_{V} \cos 2 \chi \\
\Xi_{--}=\sin ^{2} \theta_{V}\left(1+\cos \theta_{\ell}\right)^{2}, & \Xi_{+0}=-4 \sin \theta_{\ell}\left(1-\cos \theta_{\ell}\right) \sin \theta_{V} \cos \theta_{V} \cos \chi \\
\Xi_{00}=4 \sin ^{2} \theta_{\ell} \cos ^{2} \theta_{V}, & \Xi_{-0}=4 \sin \theta_{\ell}\left(1+\cos \theta_{\ell}\right) \sin \theta_{V} \cos \theta_{V} \cos \chi
\end{array}
$$

### 6.1.2 $\zeta$ parameter product basis

Though it is not immediately obvious from the form of the PDF in eq.(3.5), in fact by working through the algebraic expansion, one can from eq.(3.5) and eqs.(3.8-3.11) show that the PDF can be expanded in powers of the basis of the parameters we are trying to fit for. Specifically, in the basis:

$$
\begin{equation*}
{ }^{b} \zeta=\left\{1, R_{1}, R_{2}, R_{1}^{2}, R_{2}^{2}, R_{1} R_{2}\right\} \times\left\{1, \rho^{2}, \rho^{4}\right\} \tag{6.3}
\end{equation*}
$$

where the pre-script $b$ on $\zeta$ indexes the 18 parameter products.
Defining $\vec{y} \equiv(\vec{x}, w)$ we reexpress the PDF in this basis of $6 \times 3=18$ terms:

$$
g_{j}\left(\vec{y}_{j} ; \vec{\mu}\right)=Z_{1}(\vec{y}) R_{1}^{2} \rho^{4}+Z_{2}(\vec{y}) R_{1}^{2} \rho^{2}+\ldots+Z_{18}(\vec{y}) R_{1}^{0} R_{2}^{2} \rho^{0}
$$

More compactly:

$$
g_{j}\left(\vec{y}_{j} ; \vec{\mu}\right)=\sum_{b=1}^{18}\left[{ }^{b} \zeta(\vec{\mu})\right] Z_{b}\left(\vec{y}_{j}, w_{j}\right)
$$

The $Z_{b}$ are functional coefficients that subsume all of the angular and $w$ dependent parts of the expansion, and the ${ }^{b} \zeta$ are the 18 parameter products.

In fact, the full expansion can be shown with some work to be the exactly as shown in Section 6.1.2.1 (using the notation that the $\{++,--, 00,+-,+0,-0\}$ refer to which of the six helicity amplitude pair products the term arises from). Since addition is commutative, the order number of the terms from $1 \rightarrow 18$ is arbitrary, and the one presented below is essentially the one we used in our computer code in evaluation of these quantities (clear groupings will be seen upon inspection, however).

### 6.1.2.1 $\zeta$-basis Expansion

This is the full $\zeta$-basis expansion: (we drop the purple color on $w$ for this section)

$$
\begin{aligned}
&{ }^{1} \zeta_{R_{1}^{2} \rho^{4}}=R_{1}^{2} \rho^{4} \\
& Z_{R_{1}^{2} \rho^{4}}=\eta_{\rho^{4} R_{1}^{2}}^{++}(w) \Xi_{++}+ \eta_{\rho^{4} R_{1}^{2}}^{--}(w) \Xi_{--}+\eta_{\rho^{4}}^{+-} R_{1}^{2}(w) \Xi_{+-} \\
& \eta_{\rho^{4}}^{++} R_{1}^{2}(w)=\frac{(w-1)^{3}}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
& \eta_{\rho^{4}}^{--} R_{1}^{2}(w)=\frac{(w-1)^{3}}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
& \eta_{\rho^{4}}^{+-} R_{1}^{2}(w)=-\frac{(w-1)^{3}}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)
\end{aligned}
$$

$$
\begin{gathered}
{ }^{2} \zeta_{R_{1}^{2} \rho^{2}}=R_{1}^{2} \rho^{2} \\
Z_{R_{1}^{2} \rho^{2}}=\eta_{\rho^{2} R_{1}^{2}}^{++}(w) \Xi_{++}+\eta_{\rho^{2} R_{1}^{2}}^{--}(w) \Xi_{--}+\eta_{\rho^{2} R_{1}^{2}}^{+-}(w) \Xi_{+-} \\
\eta_{\rho^{2} R_{1}^{2}}^{++}(w)=\frac{-2(w-1)^{2}}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2} R_{1}^{2}}^{--}(w)=\frac{-2(w-1)^{2}}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2} R_{1}^{2}}^{+-}(w)=\frac{+2(w-1)^{2}}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
Z_{R_{1}^{2}}=\eta_{R_{1}^{2}}^{++}(w) \Xi_{++}+\eta_{R_{1}^{2}}^{--}(w) \Xi_{--}+\eta_{R_{1}^{2}}^{+-}(w) \Xi_{+-}
\end{gathered}
$$

$$
\begin{array}{r}
\eta_{R_{1}^{2}}^{++}(w)=\frac{(w-1)}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{R_{1}^{2}}^{--}(w)=\frac{(w-1)}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{R_{1}^{2}}^{+-}(w)=-\frac{(w-1)}{(w+1)}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)
\end{array}
$$

$$
\begin{gathered}
{ }^{4} \zeta_{R_{1} \rho^{4}}=R_{1} \rho^{4} \\
Z_{R_{1} \rho^{4}}=\eta_{\rho^{4} R_{1}}^{++}(w) \Xi_{++}+\eta_{\rho^{4} R_{1}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{4} R_{1}}^{+0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{array}{r}
\eta_{\rho^{4} R_{1}}^{++}(w)=-2(w-1)^{2} \sqrt{\frac{(w-1)}{(w+1)}}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{4} R_{1}}^{--}(w)=+2(w-1)^{2} \sqrt{\frac{(w-1)}{(w+1)}}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{4} R_{1}}^{+0}(w)=-(w-1)^{2} \sqrt{\frac{(w-1)}{(w+1)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{\rho^{4} R_{1}}^{-0}(w)=+(w-1)^{2} \sqrt{\frac{(w-1)}{(w+1)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right)
\end{array}
$$

$$
\begin{gathered}
{ }^{5} \zeta_{R_{1} \rho^{2}}=R_{1} \rho^{2} \\
Z_{R_{1} \rho^{2}}=\eta_{\rho^{2} R_{1}}^{++}(w) \Xi_{++}+\eta_{\rho^{2} R_{1}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{2} R_{1}}^{+0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{array}{r}
\eta_{\rho^{2} R_{1}}^{++}(w)=+4(w-1) \sqrt{\frac{(w-1)}{(w+1)}}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2} R_{1}}^{--}(w)=-4(w-1) \sqrt{\frac{(w-1)}{(w+1)}}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2} R_{1}}^{+0}(w)=+2(w-1) \sqrt{\frac{(w-1)}{(w+1)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{\rho^{2} R_{1}}^{-0}(w)=-2(w-1) \sqrt{\frac{(w-1)}{(w+1)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right)
\end{array}
$$

$$
\begin{gathered}
{ }^{6} \zeta_{R_{1}}=R_{1} \\
Z_{R_{1}}=\eta_{R_{1}}^{++}(w) \Xi_{++}+\eta_{R_{1}}^{+0}(w) \Xi_{+0}+\eta_{R_{1}}^{+0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{gathered}
\eta_{R_{1}}^{++}(w)=-2 \sqrt{\frac{(w-1)}{(w+1)}}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{R_{1}}^{--}(w)=+2 \sqrt{\frac{(w-1)}{(w+1)}}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{R_{1}}^{+0}(w)=-\sqrt{\frac{(w-1)}{(w+1)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{R_{1}}^{-0}(w)=+\sqrt{\frac{(w-1)}{(w+1)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
Z_{\rho^{4} R_{2}^{2}=}^{{ }^{(w)}} \eta_{\rho^{4} R_{2}^{2}(w) \Xi_{00}}^{{ }^{7} \zeta_{\rho^{4}} R_{2}^{2}}=\rho^{4} R_{2}^{2}
\end{gathered}
$$

$$
\begin{aligned}
& \eta_{\rho^{4} R_{2}^{2}}^{00}(w)=\frac{(w-1)^{4}}{(1-r)^{2}} \\
& Z_{\rho^{2} R_{2}^{2}}=\zeta_{\rho^{2} R_{2}^{2}}^{00}=\rho^{2} R_{2}^{2} \\
& R_{2}^{2}(w) \Xi_{00}
\end{aligned}
$$

$$
\eta_{\rho^{2} R_{2}^{2}}^{00}(w)=-2(w-1) \frac{(w-1)^{2}}{(1-r)^{2}}
$$

$$
\begin{array}{r}
{ }^{9} \zeta_{R_{2}^{2}}=R_{2}^{2} \\
Z_{R_{2}^{2}}=\eta_{R_{2}^{2}}^{00}(w) \Xi_{00}
\end{array}
$$

$$
\eta_{R_{2}^{2}}^{00}(w)=\frac{(w-1)^{2}}{(1-r)^{2}}
$$

$$
\begin{gathered}
{ }^{10} \zeta_{\rho^{4} R_{2}}=\rho^{4} R_{2} \\
Z_{\rho^{4} R_{2}}=\eta_{\rho^{4} R_{2}}^{00}(w) \Xi_{00}+\eta_{\rho^{4} R_{2}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{4} R_{2}}^{-0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{gathered}
\eta_{\rho^{4} R_{2}}^{00}(w)=-2(w-1)^{2} \frac{(w-1)}{(1-r)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{\rho^{4} R_{2}}^{+0}(w)=-(w-1)^{2} \frac{(w-1)}{(1-r)} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
\eta_{\rho^{4} R_{2}}^{-0}(w)=-(w-1)^{2} \frac{(w-1)}{(1-r)} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}
\end{gathered}
$$

$$
\begin{gathered}
{ }^{11} \zeta_{\rho^{2} R_{2}}=\rho^{2} R_{2} \\
Z_{\rho^{2} R_{2}}=\eta_{\rho^{2} R_{2}}^{00}(w) \Xi_{00}+\eta_{\rho^{2} R_{2}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{2} R_{2}}^{-0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{aligned}
& \eta_{\rho^{2} R_{2}}^{00}(w)=+4(w-1) \frac{(w-1)}{(1-r)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
& \eta_{\rho^{2} R_{2}}^{+0}(w)=+2(w-1) \frac{(w-1)}{(1-r)} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
& \eta_{\rho^{2} R_{2}}^{-0}(w)=+2(w-1) \frac{(w-1)}{(1-r)} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
& Z_{R_{2}}=\eta_{R_{2}}^{00}(w) \Xi_{00}+\eta_{R_{2}}^{+0}(w) \Xi_{+0}+\eta_{R_{2}}^{-0}(w) \Xi_{-0}
\end{aligned}
$$

$$
\begin{gathered}
\eta_{R_{2}}^{00}(w)=-2 \frac{(w-1)}{(1-r)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{R_{2}}^{+0}(w)=-\frac{(w-1)}{(1-r)} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
\eta_{R_{2}}^{-0}(w)=-\frac{(w-1)}{(1-r)} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
Z_{\rho^{4} R_{1} R_{2}}=\eta_{\rho^{4} R_{1} R_{2}}^{+0}{ }_{\rho^{4} R_{1} R_{2}}=\rho^{4} R_{1} R_{2} \\
\Xi_{+0}+\eta_{\rho^{4} R_{1} R_{2}}^{-0}(w) \Xi_{-0} \\
\eta_{\rho^{4} R_{1} R_{2}}^{+0}(w)=+(w-1)^{2}\left(\frac{(w-1)}{(1-r)}\right) \sqrt{\frac{(w-1)}{(1-r)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
\eta_{\rho^{4} R_{1} R_{2}}^{-0}(w)=-(w-1)^{2}\left(\frac{(w-1)}{(1-r)}\right) \sqrt{\frac{(w-1)}{(1-r)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}
\end{gathered}
$$

$$
\begin{gathered}
{ }^{14} \zeta_{\rho^{2} R_{1} R_{2}}=\rho^{2} R_{1} R_{2} \\
Z_{\rho^{2} R_{1} R_{2}}=\eta_{\rho^{2} R_{1} R_{2}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{2} R_{1} R_{2}}^{-0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{array}{r}
\eta_{\rho^{2} R_{1} R_{2}}^{+0}(w)=-2(w-1)\left(\frac{(w-1)}{(1-r)}\right) \sqrt{\frac{(w-1)}{(1-r)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
\eta_{\rho^{2} R_{1} R_{2}}^{-0}(w)=--2(w-1)\left(\frac{(w-1)}{(1-r)}\right) \sqrt{\frac{(w-1)}{(1-r)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}
\end{array}
$$

$$
\begin{gathered}
{ }^{15} \zeta_{R_{1} R_{2}}=R_{1} R_{2} \\
Z_{R_{1} R_{2}}=\eta_{R_{1} R_{2}}^{+0}(w) \Xi_{+0}+\eta_{R_{1} R_{2}}^{-0}(w) \Xi_{-0}
\end{gathered}
$$

$$
\begin{aligned}
& \eta_{R_{1} R_{2}}^{+0}(w)=+\left(\frac{(w-1)}{(1-r)}\right) \sqrt{\frac{(w-1)}{(1-r)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)} \\
& \eta_{R_{1} R_{2}}^{-0}(w)=-\left(\frac{(w-1)}{(1-r)}\right) \sqrt{\frac{(w-1)}{(1-r)}} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}
\end{aligned}
$$

$$
\begin{aligned}
& { }^{16} \zeta_{\rho^{4}}=\rho^{4} \\
& Z_{\rho^{4}}=\eta_{\rho^{4}}^{++}(w) \Xi_{++}+\eta_{\rho^{4}}^{--}(w) \Xi_{--}+\eta_{\rho^{4}}^{00}(w) \Xi_{00}+\eta_{\rho^{4}}^{+-}(w) \Xi_{+-}+\eta_{\rho^{4}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{4}}^{-0}(w) \Xi_{-0}
\end{aligned}
$$

$$
\begin{gathered}
\eta_{\rho^{4}}^{++}(w)=(w-1)^{2}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{4}}^{--}(w)=(w-1)^{2}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{4}}^{00}(w)=(w-1)^{2}\left(1+\frac{(w-1)}{(1-r)}\right)^{2} \\
\eta_{\rho^{4}}^{+-}(w)=(w-1)^{2}\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{4}}^{+0}(w)=(w-1)^{2} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{\rho^{4}}^{-0}(w)=(w-1)^{2} \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
Z_{\rho^{2}}=\eta_{\rho^{2}}^{++}(w) \Xi_{++}+\eta_{\rho^{2}}^{--}(w) \Xi_{--}+\eta_{\rho^{2}}^{00}(w) \Xi_{00}+\eta_{\rho^{2}}^{+-}(w) \Xi_{+-}+\eta_{\rho^{2}}^{+0}(w) \Xi_{+0}+\eta_{\rho_{2}^{2}}^{-0}(w) \Xi_{-0}
\end{gathered}
$$

$$
{ }^{17} \zeta_{\rho^{2}}=\rho^{2}
$$

$$
\begin{gathered}
\eta_{\rho^{2}}^{++}(w)=-2(w-1)\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2}}^{--}(w)=-2(w-1)\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2}}^{00}(w)=-2(w-1)\left(1+\frac{(w-1)}{(1-r)}\right)^{2} \\
\eta_{\rho^{2}}^{+-}(w)=-2(w-1)\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{2}}^{+0}(w)=-2(w-1) \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)\left(1+\frac{(w-1)}{(1-r)}\right)} \\
Z_{\rho^{0}}^{-0}=\eta_{\rho^{2}}^{++}(w)=-2(w-1) \sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
{ }^{18} \zeta_{\rho^{0}}=\rho^{0} \\
+\eta_{\rho^{0}}^{--}(w) \Xi_{--}+\eta_{\rho^{0}}^{00}(w) \Xi_{00}+\eta_{\rho^{0}}^{+-}(w) \Xi_{+-}+\eta_{\rho^{0}}^{+0}(w) \Xi_{+0}+\eta_{\rho^{0}}^{-0}(w) \Xi_{-0}
\end{gathered} \quad \begin{array}{r}
\eta_{\rho^{0}}^{++}(w)=\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right) \\
\eta_{\rho^{0}}^{--}(w)=\left(\frac{1-2 w r+r^{2}}{\left.(1-r)^{2}\right)}\right) \\
\eta_{\rho^{0}}^{00}(w)=\left(1+\frac{(w-1)}{(1-r)}\right)^{2} \\
\eta_{\rho^{0}}^{+0}(w)=\sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right) \\
\eta_{\rho^{0}}^{-0}(w)=\sqrt{\left(\frac{1-2 w r+r^{2}}{(1-r)^{2}}\right)}\left(1+\frac{(w-1)}{(1-r)}\right)
\end{array}
$$

Clearly there is some room for algebraic error in making this expansion, and the verifi-
cation of this form was provided by comparing the total sum numerically for a given event (i.e., set of four input kinematic variables) in the expanded PDF version, with the value from evaluating much simpler original PDF form eq.(3.5) for the same event. Only when these numbers matched exactly (to machine precision) were we assured that we had the correct expansion.

In order to be very sure of this test, we both spotchecked that the PDF forms evaluated to the same value for random events, and we also kept a running sum over all events for the same form, and verified that both sums were identical at the end of running over e.g. 100k and even 1M event files.

As examples, here is the final numerical output from running over a 10k event file with CCC parameters (the labelling of the parameter sets here follows the convention listed in Section 7.2.1 ):
(file: Rccc.rand2.B0fr.4col.10kevts)
evtnum $=10000$
totsum= 40592.554688
dgd4 $=40592.554688$
Where the "dgd4" value is from the original form of the PDF summed over all 10k events, and "totsum" is the value from the moments expansion form of the PDF.

Different file, CCC also, 100k evts:
(file: Rccc.aes254.100k)
evtnum $=100,000$
totsum $=446518.906250$
$\operatorname{dgd4}=446518.906250$

Different file, CCC also, 1M evts:

```
evtnum = 1,000,000
    totsum= 4505840.500000
    dgd4=4505840.500000
```


### 6.2 Moments With a Realistic Detector: Acceptance/Efficiency

## Effects

### 6.2.1 $\xi$ parameters

Now in general when acceptance and efficiency effects are taken into account we write that the observed distribution is of the form ${ }^{1}$ :

$$
\begin{equation*}
g^{n u m}\left(\vec{y}^{\prime} ; \vec{\mu}\right)=\int d \vec{y} g^{t h}(\vec{y} ; \vec{\mu}) \mathcal{F}\left(\vec{y}, \vec{y}^{\prime}\right) \tag{6.4}
\end{equation*}
$$

where:

- $\vec{y}=$ true kinematic variables of event (non-determinable, except for in MC)
- $\vec{y}^{\prime}=$ observed kinematic variables of event (different from the true ones due to smearing effects)
- $\mathcal{F}\left(\vec{y}, \vec{y}^{\prime}\right)$ is the combined smearing and acceptance function
- $g^{n u m}=$ observed PDF - actual number of events observed at $\vec{y}^{\prime}$
- $g^{t h}=$ properly normalized version of the theoretical PDF from eq.(3.5), so that $\int_{\vec{y}} g^{t h}(\vec{y})=1$

We may now define a new function $\mathcal{H}$ via $\mathcal{H}\left(\vec{y}, \vec{y}^{\prime}\right) \equiv \frac{\mathcal{F}\left(\vec{y}, \vec{y}^{\prime}\right)}{\epsilon(\vec{y})}$ where $\epsilon(\vec{y})$ is the actual acceptance function and all smearing is subsumed into the $\mathcal{H}$ function - that is, the combined total acceptance function $\mathcal{F}$ is broken into a product of a resolution-smearing only function $\mathcal{H}$

[^5]and a efficiency-only function $\epsilon$. (Note that the efficiency for a given event to be accepted by the detector is a function of the kinematic variables only and not the parameters since detector acceptance can only depend on the physical tracks and clusters in the detector for a given event, and cannot depend on the underlying parameters of the PDF).

Substituting $\mathcal{H}$ into eq.(6.4) we obtain:

$$
\begin{equation*}
g^{n u m}\left(\vec{y}^{\prime} ; \vec{\mu}\right)=\int d \vec{y} g^{t h}(\vec{y} ; \vec{\mu}) \epsilon(\vec{y}) \mathcal{H}\left(\vec{y}, \vec{y}^{\prime}\right) \tag{6.5}
\end{equation*}
$$

Now take the following modification of the above, where we multiply through by unity in the form of $g\left(\vec{y} ; \mu_{\mathrm{mc}}\right) / g\left(\vec{y} ; \mu_{\mathrm{mc}}\right)$ (where $\mu_{\mathrm{mc}}$ is the parameter set used to generate our MC sample):

$$
\begin{equation*}
g^{n u m}\left(\vec{y}^{\prime} ; \vec{\mu}\right)=\int d \vec{y} g^{t h}(\vec{y} ; \vec{\mu}) \epsilon(\vec{y}) \mathcal{H}\left(\vec{y}, \vec{y}^{\prime}\right) \times \frac{g\left(\vec{y} ; \mu_{\mathrm{mc}}\right)}{g\left(\vec{y} ; \mu_{\mathrm{mc}}\right)} \tag{6.6}
\end{equation*}
$$

If $\mu$ is not too different from $\mu_{\mathrm{mc}}$ (which can be adjusted iteratively by reweighting the MC to be closer to the data after the fit if needed) then the ratio $g(\vec{y} ; \mu) / g\left(\vec{y} ; \mu_{\mathrm{mc}}\right)$ will not vary much across the range where the resolution function is much different from zero. In this case we can use $\vec{y}^{\prime}$ to approximate $\vec{y}$, which allows us to pull this ratio out of the integral over $\vec{y}$ to obtain:

$$
\begin{equation*}
g^{n u m}\left(\vec{y}^{\prime} ; \vec{\mu}\right) \approx \frac{g\left(\vec{y}^{\prime} ; \mu\right)}{g\left(\vec{y}^{\prime} ; \mu_{\mathrm{mc}}\right)} \int d \vec{y} g^{t h}\left(\vec{y} ; \vec{\mu}_{\mathrm{mc}}\right) \epsilon(\vec{y}) \mathcal{H}\left(\vec{y}, \vec{y}^{\prime}\right) \tag{6.7}
\end{equation*}
$$

A further assumption we make is that the PDF itself does not vary appreciably over the width of $\mathcal{H}$ (made believable by the narrowness of the resolution plots in Figure F.1) so that we may write $\int d \vec{y} g^{\text {th }}(\vec{y} ; \vec{\mu}) \mathcal{H}\left(\vec{y}, \vec{y}^{\prime}\right) \approx g^{\text {th }}\left(\vec{y}^{\prime} ; \vec{\mu}\right)$ (this is effectively treating the $\mathcal{H}^{\prime}$ s as Dirac delta functions). Though we know this is fairly accurate, it is not completely the case, and the resulting biases in the fit will be removed via our bias map removal method of Section 8.1.

In order to do a proper (non-extended maximum likelihood) fit, we need to further normalize this smeared "PDF", so we define now the normalizing function (alternately called the "average efficiency" or the "PDF-weighted efficiency normalization factor") averaged over the detector for the entire 4-D $\vec{y}$ phase space:

$$
\begin{equation*}
\bar{\epsilon}(\vec{\mu})=\int g^{t h}(\vec{y} ; \vec{\mu}) \epsilon(\vec{y}) d \vec{y} \tag{6.8}
\end{equation*}
$$

Since we are integrating over the full kinematic variable space, clearly the efficiency normalization factor $\bar{\epsilon}$ can only be a function of the parameters $\vec{\mu}$ (in the next section we will see how we can estimate the integral as a discrete sum over events and ultimately obtain this efficiency correction factor from MC samples).

One of the major benefits of this method is that in fact the explicit form of the (complicated 4D) efficiency function $\epsilon$ is not needed (since it depends only on the kinematic variables and not on the parameters, it drops out in log-likelihood minimization) and we need only $\bar{\epsilon}$ to perform the fit to the form factors.

Thus the final properly normalized PDF that we will use in our fit is:

$$
\begin{equation*}
g^{o b s}\left(\vec{y}^{\prime} ; \vec{\mu}\right)=\frac{g_{j}^{t h}\left(\vec{y}^{\prime} ; \vec{\mu}\right) \epsilon(\vec{y})}{\bar{\epsilon}(\vec{\mu})} \tag{6.9}
\end{equation*}
$$

where $g^{\text {obs }}$ does not give the exact number of events at a given point, but the normalized distribution of observed events instead.

Define now for the $j$ 'th event in our sample (leaving out some of the arguments):

$$
\begin{equation*}
g_{j}^{\text {obs }}=\frac{g_{j}^{t h} \epsilon\left(\vec{y}_{j}\right)}{\bar{\epsilon}(\vec{\mu})} \tag{6.10}
\end{equation*}
$$

Also note that our Moments Method accounts not only for the loss of events due to geometric (detector) and tracking/PID type acceptance losses, but that due to anything resolution-
caused where some relevant variable was smeared such that it no longer falls in the acceptance after reconstruction. This is because for this method, these types of events are no different than the geometrically rejected events - that is, if the MC reasonably correctly replicates the data resolution, then it will have about the same number of events rejected in the same region and thus the Moments Method will take this into account when used to correct for the acceptance (in practice, the primary variable this occurs for is $\cos \theta_{B Y}$, whose absolute value can be smeared to be greater than unity especially due to FSR effects; once an event is accepted within the physical $\cos \theta_{B Y}$ $-1 \rightarrow 1$ cut, all four kinematic variables can be physically constructed for it.)

### 6.2.2 Expression for $\bar{\epsilon}$

A summary in words of the procedure we show below: take now the theoretical PDF eq.(3.5) in the the $\zeta$ basis form of Section 6.1.2 and substitute it into the $\bar{\epsilon}$ definition eq.(6.8). Next factor out of the integral the portion that doesn't depend on the kinematic variables, and define the resultant integrals as the $\xi_{b}$ acceptance correction functions (note here, as in Section 6.1.1, we use $\vec{x}=3$ angular kinematic variables only, and $\vec{y}=$ all 4 kinematic variables, where $w$ is included along with the angular kinematic variables):

$$
\begin{aligned}
\bar{\epsilon}(\vec{\mu}) & =\int\left[\sum_{a=1}^{6} \mathcal{H}_{a}(w ; \vec{\mu}) \Xi_{a}(\vec{x})\right] \epsilon(\vec{y}) d \vec{y} \\
& =\int\left[\sum_{b=1}^{18} Z_{b}(\vec{x}, w)\left[{ }^{b} \zeta\right]\right] \epsilon(\vec{y}) d \vec{y} \\
& =\sum_{b=1}^{18}\left[{ }^{b} \zeta\right] \int Z_{b}(\vec{y}) \epsilon(\vec{y}) d \vec{y} \\
& =\sum_{b=1}^{18}\left[{ }^{b} \zeta\right] \xi_{b}
\end{aligned}
$$

Where we have defined:

$$
\begin{equation*}
\xi_{b} \equiv \int Z_{b}(\vec{y}) \epsilon(\vec{y}) d \vec{y} \tag{6.11}
\end{equation*}
$$

These are the 18 numbers we need to obtain to do all of the detector-related, geometric and event selection acceptance corrections for this analysis.

### 6.2.3 $\xi$ Correction Factors from Sums Over Events

We will here derive a form for the $\xi_{b}$ which will allow them to be evaluated from a MC sample. We begin by discretizing the integral over space to a sum over $N$ boxes as so:

$$
\begin{equation*}
\xi_{b}=\int Z_{b}(\vec{y}) \epsilon(\vec{y}) d \vec{y} \rightarrow \sum_{i=1}^{N} Z_{b}\left(\vec{y}_{i}\right) \epsilon\left(\vec{y}_{i}\right) \Delta y \tag{6.12}
\end{equation*}
$$

- $\vec{y}_{i}$ is the value of $\vec{y}$ evaluated at the center of the box
- $b$ labels the term in the $\zeta$ expansion sum (running from 1 to 18 )
- $\Delta y$ is the 4D kinematic variable box size

Taking $N_{\text {gen }}$ to be the number of generated MC events, we may now write the occupancy of the $i^{\prime}$ th box as:

$$
\begin{equation*}
\mathcal{O}_{i}=N_{\text {gen }} g^{t h}\left(\vec{y}_{i} ; \mu_{0}\right) \epsilon\left(\vec{y}_{i}\right) \Delta y \tag{6.13}
\end{equation*}
$$

where we have evaluated the PDF $g$ also at the center of the box and at an initial value of the parameters $\mu_{0}$. This yields:

$$
\begin{equation*}
\epsilon\left(\vec{y}_{i}\right) \Delta y=\frac{\mathcal{O}_{i}}{N_{\operatorname{gen}} g^{t h}\left(\vec{y}_{i} ; 0\right)} \tag{6.14}
\end{equation*}
$$

So that as we shrink the box down to infinitesimal size such that each box contains just zero or one event, the sum eq.(6.12) becomes:

$$
\begin{equation*}
\xi_{b}\left(\vec{\mu}_{0}\right)=\sum_{i=1}^{N_{\text {gen }}} \frac{Z_{b}\left(\vec{y}_{i}\right) \mathcal{O}_{i}}{N_{\text {gen }} g^{t h}\left(\vec{y}_{i} ; \vec{\mu}_{0}\right)} \tag{6.15}
\end{equation*}
$$

And $\mathcal{O}_{i}$ has now become a binary function which is unity if the event is seen in the reconstructed MC, and zero if it is not.

Taking the box index $i$ to the event index $j$, we find this is in fact a form which we can sum over events:

$$
\begin{equation*}
\xi_{b}\left(\vec{\mu}_{0}\right)=\sum_{j=1}^{N_{\text {obs }}} \frac{Z_{b}\left(\vec{y}_{j}\right)}{N_{\text {gen }} g^{t h}\left(\vec{y}_{j} ; \vec{\mu}_{0}\right)} \tag{6.16}
\end{equation*}
$$

Where $g^{\text {th }}\left(\vec{y}_{j} ; \vec{\mu}_{0}\right)$ is the value of the PDF for the $j^{\prime}$ th event, evaluated at $(\vec{y})$ for this event, and with $\vec{\mu}$ fixed to the generated MC values $\vec{\mu}_{0}$.

These two sums are formally identical, but will have some discrepancy due to the finite statistics with which we evaluate them. In the limit of large $N_{\text {gen }}$ and vanishing $\Delta y$ they would give the same exact result, but there will be a certain amount of evaluable error (see Section 8.2.2) associated with the finiteness of the MC sample which we will use to do this "calibration" sum.

Also note that though we have written the $\xi_{b}(\vec{\mu})$ explicitly as functions of the $\vec{\mu}$ parameters which the MC sample that we evaluated the sums from was generated with, the $\xi_{b}$ should not in theory vary significantly with the parameters that the sample is generated with, and should depend only on the acceptance characteristics of the detector itself. This assumption is verified by tests described in Section 8.2.1.

### 6.2.4 Fitting to Detector-corrected Sample

We now form the log likelihood:

$$
\begin{align*}
\ln \mathcal{L}(\vec{\mu}) & =\sum_{j=1}^{N_{\text {obs }}} \ln g_{j}^{\text {obs }}  \tag{6.17}\\
& =\sum_{j=1}^{N_{\text {obs }}} \ln \left(\frac{g_{j}^{t h} \epsilon\left(\vec{y}_{j}\right)}{\bar{\epsilon}(\vec{\mu})}\right)  \tag{6.18}\\
& =\sum_{j=1}^{N_{\text {obs }}}\left[\ln g_{j}^{\text {th }}-\ln \left(\sum_{b=1}^{18}\left[{ }^{b} \zeta\right] \xi_{b}\left(\vec{\mu}_{0}\right)\right)\right]  \tag{6.19}\\
& =\sum_{j=1}^{N_{\text {obs }}}\left(\ln g^{t h}\left(\vec{y}_{j} ; \vec{\mu}\right)\right)-N_{\text {obs }} \ln \left(\sum_{b=1}^{18}\left[{ }^{b} \zeta\right] \xi_{b}\left(\vec{\mu}_{0}\right)\right) \tag{6.20}
\end{align*}
$$

Where we have dropped the $\left(\sum_{j=1}^{N_{\text {obs }}} \ln \epsilon\left(\vec{y}_{j}\right)\right)$ term in the third step as it has no dependence on the parameters, and have summed over all observed events in the second
term to give factor of $N_{\text {obs }}$ in the last step. We now simply maximize this likelihood with respect to the parameters to find their best fit values.

### 6.2.5 Self-calibration Theorem

We can show that when we fit to the same MC data used to calibrate the efficiency moments, the values returned by the fitter must be exactly the same as the MC input. This actually is not a result of the use of the moments expansion, but follows from using the MC sum over the same dataset used to evaluate the efficiency.

To see this, first recast the expression for the PDF normalization $\bar{\epsilon}$ by noting that using eq.(6.16), $\bar{\epsilon}$ can also be written as:

$$
\begin{equation*}
\bar{\epsilon}(\vec{\mu})=\sum_{b=1}^{18}\left[{ }^{b} \zeta\right] \xi_{b}=\sum_{b=1}^{18} \sum_{j=1}^{N_{\text {obs }}} \frac{Z_{b}\left(\vec{y}_{j}\right)\left[{ }^{b} \zeta\right]}{N_{\text {gen }} g\left(\vec{y}_{j} ; \vec{\mu}_{0}\right)}=\frac{1}{N_{\text {gen }}} \sum_{j=1}^{N_{\text {obs }}} \frac{g\left(\vec{y}_{j} ; \vec{\mu}\right)}{g\left(\vec{y}_{j} ; \vec{\mu}_{0}\right)} \tag{6.21}
\end{equation*}
$$

where we have collapsed the sum over the 18 products of $Z$ functions and $\zeta$ parameter combinations back into the original form of the PDF.

Now from eq.(6.18) form the log likelihood expression:

$$
\begin{equation*}
\ln \mathcal{L}(\vec{\mu})=\sum_{j=1}^{N_{\text {obs }}} \ln \left(\frac{g\left(\vec{y}_{j}, \vec{\mu}\right)}{\frac{1}{N_{g e n}} \sum_{i}^{N_{o b s}} \frac{g\left(\vec{i}_{i}, \vec{\mu}\right)}{g\left(\vec{y}_{i}, \vec{\mu}_{0}\right)}}\right) \tag{6.22}
\end{equation*}
$$

where $g$ is the un-normalized PDF, including in this case the efficiency effects, so that $g \equiv g^{\text {th }}\left(\vec{y}_{j} ; \vec{\mu}\right) \epsilon\left(\vec{y}_{j}\right), \vec{\mu}$ is the parameter set we are evaluating the likelihood at and $\vec{\mu}_{0}$ is the parameter set used to generate the data. To obtain the normalized PDF
we divided through by the normalization factor $\bar{\epsilon}$, using a Monte Carlo estimate of this average efficiency, as in eq.(6.21).

Consider now the case where we fit the same MC we use to perform the MC integral for the efficiency, so we are summing over the same set of events, and we may set $i=j$ :

$$
\begin{equation*}
\ln \mathcal{L}(\vec{\mu})=\sum_{j=1}^{N_{\text {obs }}} \ln \left(g\left(\vec{y}_{j}, \vec{\mu}\right)\right)-\sum_{j=1}^{N_{\text {obs }}} \ln \left(\frac{1}{N_{\text {gen }}} \sum_{i}^{N_{\text {obs }}} \frac{g\left(\vec{y}_{i}, \vec{\mu}\right)}{g\left(\vec{y}_{i}, \vec{\mu}_{0}\right)}\right) \tag{6.23}
\end{equation*}
$$

and since the second summand has no dependence on the index $j$,

$$
\begin{equation*}
\ln \mathcal{L}(\vec{\mu})=\sum_{j=1}^{N_{\text {obs }}} \ln \left(g\left(\vec{y}_{j}, \vec{\mu}\right)\right)-N_{\text {obs }} \ln \left(\frac{1}{N_{\text {gen }}} \sum_{i}^{N_{\text {obs }}} \frac{g\left(\vec{y}_{i}, \vec{\mu}\right)}{g\left(\vec{y}_{i}, \vec{\mu}_{0}\right)}\right) \tag{6.24}
\end{equation*}
$$

we may expand this further:

$$
\begin{equation*}
\ln \mathcal{L}(\vec{\mu})=\sum_{j=1}^{N_{\text {obs }}} \ln \left(g\left(\vec{y}_{j}, \vec{\mu}\right)\right)-N_{\text {obs }}\left[\left(\ln \frac{1}{N_{\text {gen }}}\right)+\ln \left(\sum_{i}^{N_{\text {obs }}} \frac{g\left(\vec{y}_{i}, \vec{\mu}\right)}{g\left(\vec{y}_{i}, \vec{\mu}_{0}\right)}\right)\right] \tag{6.25}
\end{equation*}
$$

and drop the $\left(-\ln \frac{1}{N_{\text {gen }}}\right)$ term, since it does not depend on the parameters $\vec{\mu}$ :

$$
\begin{equation*}
\ln \mathcal{L}(\vec{\mu})=\sum_{j=1}^{N_{\text {obs }}} \ln \left(g\left(\vec{y}_{j}, \vec{\mu}\right)\right)-N_{\text {obs }}\left[\ln \left(\sum_{i}^{N_{\text {obs }}} \frac{g\left(\vec{y}_{i}, \vec{\mu}\right)}{g\left(\vec{y}_{i}, \vec{\mu}_{0}\right)}\right)\right] \tag{6.26}
\end{equation*}
$$

Take now the derivative of this expression with respect to $\mu_{k}$, and set equal to zero to find its maximum for the $k^{t h}$ parameter:

$$
\begin{equation*}
0=\frac{d \ln \mathcal{L}(\vec{\mu})}{d \mu_{k}}=\sum_{j=1}^{N_{\text {obs }}}\left[\frac{1}{g\left(\vec{y}_{j} ; \mu\right)}-\frac{1}{g\left(\vec{y}_{j} ; \mu_{0}\right) \frac{N_{\text {gen }}}{N_{\text {obs }}} \bar{\epsilon}(\mu)}\right] \frac{d g\left(\vec{y}_{i} ; \vec{\mu}\right)}{d \mu_{k}} \tag{6.27}
\end{equation*}
$$

which, since

$$
\begin{equation*}
\bar{\epsilon}\left(\vec{\mu}_{0}\right)=\frac{N_{o b s}}{N_{\text {gen }}} \tag{6.28}
\end{equation*}
$$

(as is clear from eq.(6.21 ) ), implies that the bracketed term will go to zero at:

$$
\begin{equation*}
\mu_{k}=\mu_{k 0} \tag{6.29}
\end{equation*}
$$

That is: if moments are used to evaluate the MC integral, the success of a selfcalibration fit demonstrates that the moments actively used reproduce the PDF correctly. We should get the input parameters back exactly, up to machine precision. This is an excellent self-consistency test for this method.

### 6.3 SP4 Moments Method Validation

Initial tests on acceptance-less, and acceptance cut-affected generator-level MC samples were performed and details can be seen in Appendix D.

### 6.3.1 SP4 calibration on SP4 fit

Our next tests of the method were on SP4 MC samples (see Section 5.1.2 for more details on SP4 MC) and in this section we give fit results obtained from calibrating on SP4 and fitting back to SP4 samples. The final cuts made on these samples are all those listed in the Event Selection Chapter. Also, as this is MC, we can choose to evaluate the kinematic variables in the rest frame of the $B$; we discuss the issue of frame choice in real data in the next section.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.18 | 1.20 | .04 |
| $R_{2}$ | 0.72 | 0.70 | .04 |
| $\rho^{2}$ | 0.92 | 0.92 | .03 |

Table 6.1: Calibration on signal SP4Tot, Fit to SP4A.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.18 | 1.15 | .04 |
| $R_{2}$ | 0.72 | 0.72 | .04 |
| $\rho^{2}$ | 0.92 | 0.93 | .03 |

Table 6.2: Calibration on signal SP4Tot, Fit to SP4B.

In Tables 6.1-6.3 we are calibrating on the full SP4 signal sample and fitting directly to individual MC truth signal SP4 samples, to check the approximate error size, and correctness of the technique even with smeared samples.

The individual samples, here called A, B, and C are each are about $17.3 \mathrm{fb}^{-1}$ equivalent generic MC (correctly luminosity weighted amounts of $B^{0} \bar{B}^{0}, B^{+} B^{-}, c \bar{c}$ added together; as before there are no uds continuum background events that pass our cuts, so we need not include these events), and result in about 16.0 k signal $D^{*} \ell \nu$ events.

It is clear from the tables that the resulting fits are all within 1-2 $\sigma$, indicating the method is working to correct for the acceptance of realistic MC samples.

In Appendix E fit tests are shown for calibration and fit files with specific lepton and $D^{0}$ modes isolated.

The tests show the verification within statistical accuracy of the assumption that

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.18 | 1.19 | .04 |
| $R_{2}$ | 0.72 | 0.74 | .03 |
| $\rho^{2}$ | 0.92 | 0.91 | .03 |

Table 6.3: Calibration on signal SP4Tot, Fit to SP4C.
our calibration method works best for calibration samples that are most like the data we eventually fit.

Since in this work we are only presenting results in the cleanest $D^{0} \rightarrow K \pi e$ mode, for our final results we calibrate to and fit on events only selected in this mode.

### 6.3.2 Diamond Frame

Since the actual $B^{0}$ rest frame is not available to us in the data in this analysis because of the missing neutrino, some choice of a frame in which to evaluate the kinematic variables for a given event must be made.

Appendix F describes the older standard choice of the frame used in the previous CLEO analysis [4] and in other analyses where the $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ decay is reconstructed. This frame is normally referred to as the "Y-frame" (technically, what this yields is the average of kinematic variables given in two separate frames, but the single term "frame" is commonly used to refer to this averaging method and the kinematic variables given by it).

However, the Y-frame is not the only available frame choice for this analysis as we will see in this section, and this work is to our knowledge the first place where results
for a conceptually clearer and more logically consistent (as well as more intuitive and slightly numerically better) choice of frame have been presented.

We begin by noting that the Y -frame is not unique in the two configurations of $B$ 's that are chosen to form the kinematic variable average, and one might guess that an even better approximation to the actual $B$ trajectory could be achieved if we used the known fact that the $B^{\prime}$ s come out preferentially orthogonally to the beam axis to weight the vectors on the cone with a $\sin ^{2} \theta_{B}$ factor (this is because the electron and positron beams are effectively massless and thus completely polarized along the axis, and the resultant spin- $1 \Upsilon(4 \mathrm{~S})$ meson has spin along the axis and its decay distribution to its two spinless $B^{0}$ daughters results in their having a $\sin ^{2} \theta_{B}$ with respect to the beam axis) .

In fact, since there is nothing special about the plane which defines the $Y$-frame, and there is no particular reason the $B$ will lie in it, it can be seen that if we use this $\sin ^{2} \theta_{B}$ weighting, our maximum resolution gain will occur when we take more than just the two vectors in the Y -frame plane as potential $B$ trajectories. In general, one can take many potential $B$ vectors around the cone, weight them with the $\sin ^{2} \theta_{B}$ factor, and extract kinematic variables from these. However, in practice, what we found was that taking the vectors of the $B$ which lie in the plane orthogonal to that which defines the Y-frame, along with those in the Y-frame plane, for a total of four, all with the correct $\sin ^{2} \theta_{B}$ weighting, gave ultimately kinematic variables which had slightly better resolutions than those extracted from the Y-frame alone (without introducing any bias), and increasing to a large number of potential $B^{\prime}$ s spread around the cone gave negligible


Figure 6.1: Schematic representation of the origin of the term "Diamond Frame"
improvement beyond this while increasing the CPU time required to extract kinematic variables substantially. So we did our ultimate calculations with these four potential $B$ vectors. These form a diamond-like shape when the base of the cone is looked at face-on, hence the name 'Diamond Frame' (see Figures 6.1 and 6.2).

As an example of the improvement in the resolution, taking a sample of SP4 signal MC of about 18 k events, the resolution width (i.e., RMS of $w_{\text {reco }}-w_{\text {true }}$ for $w$ for the three frames goes from .02992 for the $\Upsilon(4 \mathrm{~S})$ frame to .02444 for the Y-frame (improvement of about $22 \%$ ) to .02420 for the diamond frame (improvement of just $1 \%$ ), with similar sizes for the other kinematic variables. Not a very large improvement, but we ought in any case use this frame since the increase in CPU time to include the extra information involved is minor (for at least the diamond frame which has just four $B$ configurations to calculate the kinematic variables in) and there is no reason not to take advantage of it.


Figure 6.2: Side view of $\cos \theta_{B Y}$ reconstruction. The points at $\phi_{B Y}=0$ and $\pi$ are in the $D^{*}-\ell$ plane. The points at $\pm \pi / 2$ are out of the plane.

### 6.4 Conceptual Description of Background Subtraction Method

Now we take the final analysis step of dealing with the remaining backgrounds that cannot be eliminated by other cuts from the data sample.

As it is quite difficult to find a four dimensional PDF to fit the background shape adequately, we embarked on trying different methods to subtract out our background (which amounts to about $15-25 \%$ of the total events under the $\delta m$ peak, depending on the $D^{0}$ decay mode). One we eventually found which had good success in tests is a method we dubbed the "Direct Unbinned Background Subtraction" (DUBS) method, in which we take our entire event sample, and after all final cuts are applied, subtract off a correctly luminosity weighted amount of MC truth sample of background, in order to find a final sample that is close to the signal distribution in the four relevant kinematic variables for our PDF. We then fit this leftover sample of events - which should be fairly representative of the true signal - directly to the signal PDF, and extract the fit parameters.

In the simplest form, what we are doing formally is maximizing this pseudolikelihood:

$$
\begin{equation*}
\ln \mathcal{L}=\sum_{i}^{\text {all }} \ln g_{s}-\eta \sum_{j}^{\text {bkgd }} \ln g_{s} \tag{6.30}
\end{equation*}
$$

where the factor $\eta=\frac{\mathcal{L}_{\text {data }}}{\mathcal{L}_{m c}}$ does the correct luminosity weighting of the sample (e.g. if there is twice as much MC as data, then each MC background point will only count half as much in the $\log$ likelihood sum). This is referred to as a "pseudo-likelihood" because true likelihoods are only products of the PDF evaluated over events, never ratios as we are taking here. We are evaluating here the signal PDF $g_{s}$ over all reconstructed data events
after final cuts in the first sum, and over the MC truth background events in the second (subtracted) sum. Upon exponentiating, this pseudo-likelihood can also be written as:

$$
\begin{equation*}
\mathcal{L}=\frac{\left[\prod_{i}^{\text {all }} g_{s}(i)\right]}{\left[\prod_{j}^{M C b k g d} g_{s}(j)\right]^{\eta}}=\frac{\left[\prod_{i}^{\text {data signal }} g_{s}(i) \prod_{k}^{\text {data } b k g d} g_{s}(k)\right]}{\left[\prod_{j}^{\text {MCbkgd }} g_{s}(j)\right]^{\eta}} \simeq \prod_{i}^{\text {data signal }} g_{s}(i) \tag{6.31}
\end{equation*}
$$

where the last equality will only hold insofar as the MC background sample distribution represents that of the data. There are several tests we can do of this, and the uncertainty in this statement of equivalence is indeed one of the major sources of systematic error in this method.

In actuality, as we will see in Chapter 7 the ratio between number of events in the $\delta m$ sideband region (which is almost completely combinatoric background) and signal window, $N_{\mathrm{sb} / \mathrm{sw}}^{\mathrm{evts}}$ is not identical in MC and data (it is higher in data, reflecting a smaller $D^{*} \ell \nu \mathrm{BR}$ than was put into the MC production files). Because of this we take a more sophisticated form in which we treat the peaking and combinatoric backgrounds separately. Because the ratio of the number of events in the sideband region should correspond exactly to a luminosity ratio and does not depend on the precise number entered for the $D^{*} \ell \nu$ signal MC BR as entered in the MC production files (which was based on previous PDG averages, and which we know now from the current average of world measurements is about $20 \%$ too high), in theory we could take an absolute ratio of the number of events in the sideband region as the weighting factor for the combinatorial background piece (we call this factor $\lambda_{\mathrm{sb}}=N_{\mathrm{sw}}^{\text {dataevts }} / N_{\mathrm{sb}}^{\mathrm{MCevts}}$ ).

For the $K \pi e$ mode, we find this ratio is $\sim \lambda_{\mathrm{sb}}=6.3 \%$, which is about the same
level as it is found in the two other BaBar analyses that have measured it ([36],[42]).
As for the peaking background piece (i.e., the sum of $D^{* *}$, nonresonant $B^{0} \rightarrow D^{*} \pi \ell \nu$ decays - good $D^{* \prime}$ s with badly reconstructd leptons): since very careful fits to the $\cos \theta_{B Y}$ distributions for the same data set have been performed by the BABAR analyses for the $D^{*} e \nu$ branching ratio and $V_{c b}$ yielding precise numbers for the amount of estimated peaking background in all four of the measured $D^{0}$ decay modes, we take our estimates for the amount of peaking background per $D^{0}$ mode from their fit values in [36], and label these $\lambda_{\mathrm{sw}, \mathrm{d}}$ (where $s w$ indicates the $\delta m$ signal window). These factors multiply the luminosity ratios (which are taken to be the ratio of number of events) in the signal window region for each $D^{0}$ decay mode, labelled $\eta_{\mathrm{sw}, \mathrm{d}}$. For the $K \pi e$ mode, we use $\lambda_{\mathrm{sw}}=10.9 \%$.

The pseudo-likelihood we actually end up evaluating in the case of data then is:

$$
\begin{equation*}
\ln \mathcal{L}=\sum_{i}^{\text {all }} \ln g_{s}-\sum_{d=D^{0}{ }_{\text {mode }}}^{4}\left\{\eta_{\mathrm{sw}, \mathrm{~d}} \times \lambda_{\mathrm{sw}, \mathrm{~d}} \sum_{j}^{\text {pking bkgd }} \ln g_{s}+\lambda_{\mathrm{sb}, \mathrm{~d}}^{\text {combinatorial }} \sum_{j}^{\text {bkgd }} \ln g_{s}\right\} \tag{6.32}
\end{equation*}
$$

Further, there is one technical point that needs to be mentioned: the issue of evaluating the PDF in regions of the phase space where the PDF is very close to zero, leading to very large spurious values of the log-likelihood. The way we chose to deal with this issue was through a procedure called " $\chi$-smearing", which we discuss in more detail in Appendix G.

### 6.5 Full MC Validation of DUBS Method

Initial successful DUBS tests on detectorless EvtGen samples were performed and details can be seen in Appendix H. We present below some results of the method on SP4 MC samples.

### 6.5.1 SP4 calibration on SP4 fit

In this section we give fit results obtained from DUBS subtraction tests on SP4 samples. The tests are successful in the end, with the fit results all falling within 1-2 $\sigma$.

As normal, the final cuts made are all those listed in the Event Selection Chapter. The individual samples, here called A, B, and C are each are about $17.3 \mathrm{fb}^{-1}$-equivalent generic SP4 (again see Section 5.1.2 for details).

In the series of tables shown in Table 6.4, we do a background subtraction, and can observe the results. What we see is that the majority of these fits are also within one sigma, but a few leak out into two sigma, and two are at three sigma - this is due to the underestimation of the error discussed in detail in Section 8.3, and these fit within one to two sigma when the actual error including DUBS contribution is calculated for these fits (in brief, the physical basis of this underestimation is that the shape of the likelihood distribution has not been fully accounted for by Minuit because we used a pseudo-likelihood vs. a true likelihood function).

| Parameter | Input | Fit | Error | Parameter | Input | Fit | Error |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.18 | 1.19 | .04 | $R_{1}$ | 1.18 | 1.18 | .04 |
| $R_{2}$ | 0.72 | 0.73 | .03 | $R_{2}$ | 0.72 | 0.75 | .03 |
| $\rho^{2}$ | 0.92 | 0.84 | .03 | $\rho^{2}$ | 0.92 | 0.83 | .03 |
| Parameter | Input | Fit | Error | Parameter | Input | Fit | Error |
| $R_{1}$ | 1.18 | 1.24 | .04 | $R_{1}$ | 1.18 | 1.22 | .04 |
| $R_{2}$ | 0.72 | 0.70 | .03 | $R_{2}$ | 0.72 | 0.72 | .03 |
| $\rho^{2}$ | 0.92 | 0.90 | .03 | $\rho^{2}$ | 0.92 | 0.89 | .03 |
| Parameter | Input | Fit | Error | Parameter | Input | Fit | Error |
| $R_{1}$ | 1.18 | 1.16 | .04 | $R_{1}$ | 1.18 | 1.15 | .04 |
| $R_{2}$ | 0.72 | 0.70 | .03 | $R_{2}$ | 0.72 | 0.71 | .03 |
| $\rho^{2}$ | 0.92 | 0.90 | .03 | $\rho^{2}$ | 0.92 | 0.90 | .03 |

Table 6.4: Calibration on signal SP4Tot, Fits, from left to right, of: first row: A-B, A-C, second row: B-A, B-C, third row: C-A, C-B (all in form: [ SP4X Reco cut - (scale factor) * SP4Y MC truth background], where $\mathrm{X}, \mathrm{Y}=$ one of $\mathrm{A}, \mathrm{B}, \mathrm{C}$ ). Given are the input values, the output fit values, and the statistical error as returned by Minuit.

### 6.6 Bias Map Method for Correcting for Resolution

While we have shown how we correct for acceptance with the moments method in earlier sections of this Chapter, as we can see from Appendix I, one final step remains to derive a result from a fit to the data: correction for the imperfect resolution of the measured kinematic variables. Because we do not have a four dimensional resolution function with which to convolve the theoretical PDF and do the fit, we must find another way to correct for this effect after the fact. We term the method we have chosen to implement the "Bias Map Correction" technique, and it entails first taking the fit of MC truth data with MC truth calibration, then comparing this to a fit of the same events but with MC reconstructed data with MC reconstructed calibration, and then finally, taking the difference in the results for the three form factor parameters. These differences are then
taken to be the overall final corrections due to the unknown resolution function, and they are applied back to the fits to the data to deconvolute the output numbers and get back to the original results that would have been obtained on data if we had been able to take data without any resolution effects.

This method takes into account the total resolution, which is composed of both the detector smearing and frame smearing effects (described in Appendix F), though the latter is by far the dominant effect.

The derivation of the actual implementation naturally precedes the discussion of the error from this source and is thus described in Section 8.1.1.

### 6.6.1 Bias Map Correction Error

Though we will see that because of our relatively good resolution, the ultimate biases we obtain are relatively small, there is an error still associated with this method of assuming initially that the PDF describes the data perfectly and correcting for its imperfectness later, and we will describe this in Section 8.1.1.1. Also, we use the weighted Diamond Frame as described in Section 6.3.2 for evaluation of the kinematic variables, which gives the best overall resolution for each kinematic variable compared to other possible frames we have tested (e.g. the Y-frame).

### 6.6.1.1 Bias Map Error is statistical

If we had an actual four dimensional resolution function with which to convolute the PDF and do the fit, we would find that the statistical error for the fit would
increase (because the result of using the resolution function is to broaden the likelihood distribution and thus increase the one $\sigma$ error associated with this distribution). But since we are not able to use a resolution function and must make the correction as well as estimate the error after the fact, the error that we derive properly goes into the statistical category, and is not a systematic error.

Note that the error we are referring to here is intrinsic to this method of correcting for the bias after the fact, and occurs because the shape of the likelihood function is improperly described when a resolution function is not included in the PDF.

### 6.6.2 Numerical Values of Bias Map Correction

The bias map method only has an actual effect when the true value is different from that of the generated one in the MC that the acceptance correction was done with (see Section I and specifically eq.(I.29) for this). Thus all the MC we have which has been generated with the CLEO parameters cannot be used to model the Bias Map Correction.

However, the parameters will indeed in general be different in MC and data, and thus we will need to apply a bias map correction there to get back to the true parameters.

### 6.7 Fit results

Now that we have described the entire fitting and Bias Map Correction procedure, we give the results on our data sample:

$$
\begin{align*}
R_{1} & =1.34 \pm 0.057 \\
R_{2} & =0.91 \pm 0.044  \tag{6.33}\\
\rho^{2} & =0.77 \pm 0.039
\end{align*}
$$

where the error is statistical only (including the increase factor from the Bias Map Correction, as well as the DUBS irreducible data statistical piece discussed in 8.3). The pure statistical error from Minuit is ( $0.048,0.040,0.035$ ).

Other errors including MC statistical and systematic are discussed in Chapter 8.
The errors are are highly correlated, so that in order facilitate meaningful crossexperiment comparison, it is worthwhile to give the error matrix for the full statistical error (including here the Monte Carlo statistical error we will show later). It is:

$$
\begin{array}{ccc}
0.00392 & -0.00181 & 0.001281 \\
-0.00181 & 0.002381 & -0.00165 \\
0.001281 & -0.00165 & 0.001847
\end{array}
$$

The correlations are:

$$
\begin{gather*}
\rho_{R_{1}-R_{2}}=-0.59 \\
\rho_{R_{1}-\rho^{2}}=+0.48  \tag{6.34}\\
\rho_{R_{2}-\rho^{2}}=-0.79
\end{gather*}
$$

### 6.8 Goodness of Fit Evaluation Method

Though we are doing a unbinned maximum likelihood fit to obtain the parameters, we may evaluate the GOF (Goodness of Fit) of our final parameters by doing a simple binned $\chi^{2}$ test (see [44] for a statistical justification of this method of using a binned $\chi^{2}$ to evaluate GOF after doing a unbinned maximum likelihood fit to obtain the parameters from a dataset).

### 6.8.1 GOF on Reweighted Samples

We did several stages of tests of this method with both EvtGen and SP4, whose results are shown in Appendix J. The success of these tests gave us confidence that when we moved to data and ran the fit, we would be able to reweight the MC with the extracted parameters and to obtain a $\frac{\chi^{2}}{\text { dof }}$ that indicated the GOF. That is, we extracted parameters, reweighted each point in the MC by a number which had as the numerator the PDF with these parameters, and as denominator the PDF with the parameters the MC was generated with. We then took the $\frac{\chi^{2}}{\text { dof }}$ between the reweighted MC and the data.

### 6.8.2 GOF Results on our Dataset

We consider here two forms of goodness of fit tests:

- Five 1D distributions ( the projections $w, \cos \theta_{\ell}, \cos \theta_{V}$, and $\chi$ plus the distribution of CM lepton momentum $p_{l}^{*}$ )
- Binned $\chi^{2}$ based on $6 \times 6 \times 6 \times 6$ (a total of 1296) bins. For a comparison we also
give the unweighted results which correponds to the goodness-of-fit values using the CLEO parameters [4].

Figures 6.3-6.7 compare the projections in the kinematic variables and $p_{l}^{*}$ between the MC and the data. The top plots in each figure shows the result for the CLEO parameters and while the bottom gives that obtained by re-weighting by our parameters. The left hand plots show a direct overlay of the distributions with red points for the data and green for the MC. The right hand plots show the ratio of data to MC. The fit is to a constant from which we read off the $\chi^{2}$.

Table 6.5 gives the $\chi^{2}$ for fitting each plot to a constant. The constant is always consistent with unity as it must be. In every case the agreement between MC and data improves when we use our result - sometimes substantially.

| variable | $\chi^{2}$ (prob.) CLEO | $\chi^{2}$ (prob.) ours |
| :---: | :---: | :---: |
| $w$ | $14.08(0.120)$ | $13.69(0.134)$ |
| $\cos \theta_{\ell}$ | $17.12(0.047)$ | $7.626(0.572)$ |
| $\cos \theta_{V}$ | $40.93(0.000)$ | $17.81(0.0374)$ |
| $\chi$ | $8.133(0.521)$ | $7.082(0.623)$ |
| $p_{l}^{*}$ | $17.88(0.037)$ | $7.316(0.604)$ |

Table 6.5: $\chi^{2}$ and $\chi^{2}$-probability for kinematic variable $1 D$ projections and lepton momentum; the number of bins in these plots is 10 . The number of degrees-of-freedom is 9 , since we have forced the normalization to be equal between data and MC.

The value of the $4 D$ binned $\chi^{2}$ goes from 1274.04 with CLEO parameters to 1232.23 with our parameters - an improvement of $\sim 42$ units of $\chi^{2}$. The $\chi^{2}$ per bin is slightly smaller than unity. It is not really proper to interpret this as a probability as there are about 200 empty bins which contribute nothing to $\chi^{2}$. Nonetheless it is clear that the
reweighted MC follows the distribution of the data quite well.


Figure 6.3: Data/MC comparison for $w$-distribution (for this and all following plots, red points are data, and green are MC). Top uses CLEO parameters and bottom is reweighted to our parameters.


Figure 6.4: Data/MC comparison for $\cos \theta_{l}$-distribution. Top use CLEO parameters and bottom is re-weighted to our parameters.


Figure 6.5: Data/MC comparison for $\cos \theta_{V}$-distribution. Top uses CLEO parameters and bottom is re-weighted to our parameters.


Figure 6.6: Data/MC comparison for $\chi$-distribution. Top uses CLEO parameters and bottom is re-weighted to our parameters.


Figure 6.7: Data/MC comparison for $p_{l}^{*}$-distribution. Top uses CLEO parameters and bottom is re-weighted to our parameters.

## Chapter 7

## Relevant MC and Data Plots

We now show some comparisons of MC and Data in both the kinematic variables and other relevant quantities on which we cut. All plots unless otherwise specified are from Release 10 Data and SP4.

Many other tests and comparisons were done, some quite interesting, but as they are not all central to the central thrust of the analysis, we have chosen to place many of these in Appendix K.

### 7.1 Signal vs. Backgrounds in several basic variables

First from the MC we show some of the basic variables which we work with and cut on: $\delta m$ over the entire range (Figure 7.1) and in the signal window (Figure 7.2), and $\cos \theta_{B Y}$ in Figure 7.3.


Figure 7.1: This plot shows the contribution of the signal and the different background types to the deltaM distribution.


Figure 7.2: This plot shows the contribution of the signal and the different background types to the deltaM distribution in the signal window.


Figure 7.3: This plot shows the contribution of the signal and the different background types to the $\cos \theta_{B Y}$ distribution.

### 7.2 MC vs. Data Plots and Yields

The legend in all following plots is: solid red is Data and dashed blue is MC. As noted in the captions, some are in the $\delta m$ peaking signal window $(\delta m=.1429 \rightarrow .1479 \mathrm{GeV})$, some in the $\delta m$ high sideband region $(\delta m=.150 \rightarrow .165 \mathrm{GeV})$, which is selected in that range to be almost pure combinatorial background, not including any significant percentage of the peaking signal and backgrounds, and avoiding any edge effects in the $\delta m$ spectrum downturn at the high end of the range. Thus, in the signal window both true signal and background (both peaking and combinatorial) are included in the distribu-

| Source | $\delta m$ Signal Window | Sideband Region |
| :---: | :---: | :---: |
| Data | 16.1 k | 2.2 k |
| SP4 MC | 57.6 k | 6.6 k |
| Data/MC | 0.280 | 0.334 |

Table 7.1: Ratios of number of event yields in Release 10 data over those in SP4 MC, in signal window and sideband regions.
tions, where in the sideband region the comparison is between almost pure combinatorial background (wrongly reconstructed $D^{*}$ ) for each displayed distribution.

When showing the comparisons in the plots, we normalize to the same number of events in each case. For the Release 10 Data and SP4 MC in the plots, this means we scale the MC by about half down to the number of Data events, as can be seen in Table 7.1.

There is a clear discrepancy here between the data and MC in the ratio of number of events in the signal window and sideband region, indicating a difference in efficiency in reconstructing signal vs. background between the two, which we take account of when subtracting the background (see Section 6.4 for details on how we do this).

Several plots follow: for the first few, we show plots at a large scale, e.g. the first plot shows the four kinematic variable distributions for signal and MC , the second shows the ratio of Data histograms divided by the MC histograms, with the horizontal line indicating a value of unity. The errors shown are purely statistical. These include the statistical uncertainty in the number of MC and Data events that are produced from the underlying distribution (which enters the normalization) and the bin by bin uncertainty. For perfect agreement between Data and MC we would have the line at unity going through most of the error bars. The $\frac{\chi^{2}}{d o f}$ shown is for the difference from the line at unity.

The one-dimensional kinematic variables are shown in projection, i.e., summed over the other three kinematic variables. Later plots also show basic quantity comparisons, such as slow pion momentum, lepton momentum, and $\cos \theta_{B Y}$.

In all kinematic variable plots following after the first two (which have been placed in Appendix K), the projection plots for MC and data are shown on the left hand side of the page, and the ratioplots on the right hand side. To make comparisons, plots are broken up by lepton mode and $D^{0}$ decay mode.

We also show some two-dimensional correlation plots (there are six between the four kinematic variables) projected over the other two kinematic variables - this gives also some sense of agreement between data and MC that cannot be seen in the 1D projection plots. These are also found in Appendix K.

Because different features can be seen in each type, we present both box and lego plots. Further, we present ratio and difference plots where the number of events for data and MC is taken for each two-dimensional square. We may then calculate a $\chi^{2}$ bin by bin, and we present this also, to see what regions contribute most to the $\chi^{2}$ total for that plot. Finally, we also present the $\chi^{2}$ contributions in a bin by bin text format, for numerical comparison.

Note that the CLEO-measured form factors are used in the produced MC, so that we do not in general expect exact agreement between the displayed data and MC histograms in the signal region, because there will be difference due to how different the measured form factors turn out to be from the prior CLEO form factors. This is not
true of the background though, there the input form factors do not contribute (in any direct way) and it is instead the MC validity of the simulation of the contribution to the various pieces of the Data background that is being tested since we are adding up many background categories at once.

### 7.2.1 EvtGen Plots

There are also some EvtGen plots towards the end of the collection of plots where we have control over the parameters put into the MC: Figures 7.7-7.9 show 1D projections of the Kinematic Variables from three different sets of 100k generated EvtGen datasets. The legend here is as follows: "XYZ" refers to the value of ( $R_{1} ; R_{2} ; \rho^{2}$ ) respectively. The letter codes are these: $\mathrm{A}=0.50, \mathrm{~B}=0.75, \mathrm{C}=1.00, \mathrm{D}=1.25, \mathrm{E}=1.50$, and L are the CLEO values of (1.18; $0.72 ; 0.92$ ). So e.g. for the set labelled "LLX" (meaning LLA to LLE) we have $R_{1}=1.18, R_{2}=0.72$, and $\rho^{2}$ varying from 0.50 to 1.50 .

We denote this "Standard Form Factor Parameter Notation".
Note that these plots do not have acceptance included, so they are expected to differ from the SP4 plots, yet some idea of the behavior of the kinematic variables and more basic underlying variables due to variation in the parameters can be ascertained from these plots.

Similarly, Figures K.19-K. 20 show the slow pion momentum distribution for the same EvtGen parameter-varying datasets.

Finally, Figures K.23-K. 24 show the lepton momentum distribution for the same

EvtGen parameter-varying datasets.

### 7.3 Relevant Full MC (SP4) vs. Data and EvtGen Plots

### 7.3.1 Primary Signal Region SP4 vs. Data Plots



Figure 7.4: Kinematic Variables in the signal window for both leptons. Upper left is $w$, upper right $\cos \theta_{\ell}$, lower left $\cos \theta_{V}$, and lower right is $\chi$. Data is solid red and MC is dashed blue. This is the pattern that all following plots will follow, though they will be smaller.


Figure 7.5: Kinematic Variables in the sideband region for both leptons.


Figure 7.6: Kinematic Variable ratios in the sideband region for both leptons, the $\frac{\chi^{2}}{\text { d.o.f. }}$ of $0.92,1.32,1.03$, and 1.57 (for $w, \cos \theta_{\ell}, \cos \theta_{V}, \chi$ ) indicate the amount of agreement in the combinatorial background between Data and MC.

### 7.3.2 EvtGen kinematic variable Plots with varying parameters



Figure 7.7: Kinematic Variables for EvtGen produced files, the legend uses the shorthand notation described in the text. These are the XLL plots.


Figure 7.8: These are the LXL plots.


Figure 7.9: These are the LLX plots.

### 7.3.3 Lepton momentum in SP4 vs. Data and EvtGen



Figure 7.10: Left: Lepton momentum in the signal window, Right: ratio of data to MC of these histograms. The clear upward trend in the ratio indicates that the form factors in the data are not exactly those in the MC, since the form factors affect the lepton momentum spectrum significantly as can be seen in Figures K. 23 - K. 24 .


Figure 7.11: Lepton momentum in the sideband region, here it is indicated that the MC is doing much better at simulating the background lepton momentum distribution.

## Chapter 8

## Systematic Error Estimation

We describe here our approaches to evaluating each component of the systematic error. The actual numerical results for the errors are listed in Table 8.5 at the end of this chapter.

First we give a brief outline of the types of errors described in this chapter to give a structure to the rest of what follows.

There are four major types of error:

- The first is the standard statistical error that Minuit yields when fitting to a sample, where Minuit assumes that the shape of the likelihood distribution it is given to maximize is correct.
- The second is an addition to the statistical error due to the fact that in fact the likelihood distribution that we are feeding to Minuit is not actually correct due to the fact that we have not taken resolution into the fit (some of the formulae from Sec-
tion 6.6 will be required to make the estimates of this error). In the end the scaling up of the Minuit-quoted statistical error is at about the $20 \%$ level.
- The third are MC statistical errors: errors due to limited MC statistics and which can be reduced by larger MC event samples (these require some specialized techniques for treatment, given the somewhat unusual nature method of our analysis approach).
- The last are the actual systematic errors due to misreconstruction of tracks and other standard detector-related issues (these errors are fairly typically treated).

We combine the statistical and additional statistical error to give one total final statistical error, but keep the last two categories as separate until the end so we can observe what the difference contributions to the overall error from these various sources are.

We will see that for the 2000-2002 $78 \mathrm{fb}^{-1} K \pi e$ sample we are statistically limited, but the systematic errors are not far behind in size (and the MC statistical errors are of third and least significant import), and when the analysis is extended to all $D^{0}$ modes and muons, it is likely that it will become either systematic-limited, or close to this.

### 8.1 Kinematic Smearing

### 8.1.1 Bias Map Correction and Errors with Multiple Parameters

As we have discussed in Section 6.6, since we are not doing a convolution of the

PDF with a resolution model (which is very difficult to find an accurate form for in four dimensions), we take into account the effect of neglecting resolution by the "bias map" method, whereby we evaluate the bias using MC. The effect is not to just shift the central values, but errors and correlations are also changed, which we must take into account.

We may describe the method as follows: in general, there will exist functions $C_{\alpha}(\mu)$ which take into account the combined kinematic plus detector smearing and map the true parameters $\mu$ to the observed parameters $\xi_{\alpha}$, as so:

$$
\begin{equation*}
\xi_{\alpha}=C_{\alpha}(\mu) \tag{8.1}
\end{equation*}
$$

where for the time being we assume the mapping functions $C_{\alpha}$ are perfectly known (in reality they will suffer from MC statistics errors and systematic errors which we evaluate further on.)

In the case where the shifts are not large we may use a Taylor expansion in a region around the point at which we have made the $\mathrm{MC}\left(\mu_{0}\right)$ to evaluate the $C_{\alpha}$, i.e.:

$$
\begin{equation*}
\xi_{\alpha}=C_{\alpha}\left(\mu_{0}\right)+\left.\sum \frac{\partial C_{\alpha}}{\partial \mu_{\beta}}\right|_{\mu_{0}} \Delta \mu_{\beta} \tag{8.2}
\end{equation*}
$$

or:

$$
\begin{equation*}
\Delta \xi_{\alpha}=\left.\sum \frac{\partial C_{\alpha}}{\partial \mu_{\beta}}\right|_{\mu_{0}} \Delta \mu_{\beta} \tag{8.3}
\end{equation*}
$$

where $\Delta \xi_{\alpha}=\xi_{\alpha}-C_{\alpha}\left(\mu_{0}\right)$ is the deviation of the observed value in the data from the observed value for a nearby MC sample.

Defining $\left.E_{\alpha \beta}^{-1} \equiv \frac{\partial C_{\alpha}}{\partial \mu_{\beta}}\right|_{\mu_{0}}$ we may solve for the $\Delta \mu^{\prime}$ s as follows:

$$
\begin{equation*}
\Delta \mu_{\alpha}=\sum E_{\alpha \beta} \Delta \xi_{\beta} \tag{8.4}
\end{equation*}
$$

The correct answer for the distortion introduced by neglecting the resolution is then just

$$
\begin{equation*}
\mu^{a n s}=\mu^{M C}+\Delta \mu \tag{8.5}
\end{equation*}
$$

(here $\mu^{M C}=\mu_{0}$ ).

### 8.1.1.1 Error Matrix Expansion Factor

Moving now to the modification of the original error matrix, we begin by employing the variation method (also used and described in more detail in Section 8.2.2):

$$
\begin{equation*}
\delta \mu_{\alpha}^{\text {ans }}=\delta\left(\Delta \mu_{\alpha}\right)=\sum E_{\alpha \beta} \delta\left(\Delta \xi_{\beta}\right)=\sum E_{\alpha \beta} \delta \xi_{\beta} \tag{8.6}
\end{equation*}
$$

The error matrix we need is:

$$
\begin{equation*}
\left\langle\delta \mu_{\alpha}^{a n s} \delta \mu_{\alpha^{\prime}}^{a n s}\right\rangle=\sum_{\beta \beta^{\prime}} E_{\alpha \beta} E_{\alpha^{\prime} \beta^{\prime}}\left\langle\delta \xi_{\beta} \delta \xi_{\beta^{\prime}}\right\rangle=\sum_{\beta \beta^{\prime}} E_{\alpha \beta} E_{\alpha^{\prime} \beta^{\prime}} \mathbf{e}_{\beta \beta^{\prime}} \tag{8.7}
\end{equation*}
$$

where $\mathbf{e}$ is just the error matrix Minuit gives us when we run the fit.
Since the $C_{\alpha}$ function derivatives are what feed into the $E$ matrices, we need to evaluate these. We do this by reweighting the MC to a point not too far from what the answer turns out to be. To evaluate the discrete derivates, we also need to reweight to a few nearby points shifted by a small amount (so that a linear expansion is still valid) in each variable.

### 8.1.2 How resolution enters the calculation

The resolution enters into this calculation starting at eq. (8.3): $\xi$ is the measured parameter while $C$ is the function that maps the true parameter $\mu$ to $\xi$. If there were no smearing, the map would be exactly linear and diagonal, so that $d C / d \mu$ would be a straight line function. however, if this derivative goes smaller than unity, then the variation of the measured $\xi$ goes smaller than a given variation of $\mu$-in effect we are "losing" information in the fit. This makes the inverse of the derivative in eq. (8.4), $E$, larger than unity, which is exactly what we find when we actually calculate these derivatives (note that these are all actually $3 \times 3$ matrices in our case as in eq. (8.7), so the actual results involve correlations, but we are only giving a one dimensional flavor of how the resolution enters, here).

So when there is actual resolution, as happens in the real world, then it enters through the $E$ matrix, which inflates the measured Minuit error matrix $e$ of eq. (8.7).

### 8.1.2.1 Derivatives without multiple data samples

To avoid the need for multiple data samples generated at slightly different values of the parameters, we may compute the needed derivatives ( $\frac{\partial C_{\alpha}}{\partial \mu_{\beta}}$ ) using a re-weighting technique.

In the linear approximation we have:

$$
\begin{equation*}
\Delta \mu_{\alpha}^{o b s}=\mu_{\alpha}^{o b s}-\mu_{\alpha}^{G}=\left.\sum \mathbf{e}_{\alpha \alpha^{\prime}} \frac{\partial \ln L}{\partial \mu_{\alpha^{\prime}}}\right|_{\mu_{G}} \tag{8.8}
\end{equation*}
$$

where $\mu_{G}$ is the guess around which we expand in order to solve for $\ln L_{\max }$ in the linear approximation. We take a re-weighted $\ln L$ defined as follows:

$$
\begin{equation*}
\ln L(\mu)=\sum_{M C} \frac{F\left(x_{i} ; \mu\right)}{F\left(x_{i} ; \mu_{M C}\right)} \ln F\left(x_{i} ; \mu_{M C}\right) \tag{8.9}
\end{equation*}
$$

where $i \in M C$.
The derivatives we want are

$$
\begin{equation*}
\frac{\partial C_{\alpha}}{\partial \mu_{\beta}^{\prime}}=\frac{\partial \Delta \mu_{\alpha}^{o b s}}{\partial \mu_{\beta}^{\prime}}=\left.\left.\sum \mathbf{e}_{\alpha \alpha^{\prime}} \sum_{M C} \frac{1}{F\left(x_{i} ; \mu_{M C}\right)} \frac{\partial F\left(x_{i} ; \mu\right)}{\partial \mu_{\beta}^{\prime}}\right|_{\mu_{M C}} \frac{\partial \ln F\left(x_{i} ; \mu_{M C}\right)}{\partial \mu_{\alpha^{\prime}}}\right|_{\mu_{G}} \tag{8.10}
\end{equation*}
$$

$$
\begin{equation*}
=\left.\left.\sum \mathbf{e}_{\alpha \alpha^{\prime}} \sum_{M C}^{\sum} \frac{\partial \ln F\left(x_{i} ; \mu^{\prime}\right)}{\partial \mu_{\beta}^{\prime}}\right|_{\mu_{M C}} \frac{\partial \ln F\left(x_{i} ; \mu\right)}{\partial \mu_{\alpha^{\prime}}}\right|_{\mu_{G}} \tag{8.11}
\end{equation*}
$$

where we have neglected $\partial e / \partial \mu$ terms which are of higher order if $\mu_{G}$ is well chosen.
Eq.(8.11) can be evaluated by summing over the MC sample used for determining the biases using the derivatives as described in the previous section.

### 8.1.3 Results From SP4

We take two separate 13k SP4 pure signal samples representing a calibration and data file, and run the fitter to first find:

## Minuit Error Matrix:

$$
\begin{array}{ccr}
0.002379 & -0.001367 & 0.0007875 \\
-0.001367 & 0.001488 & -0.0008942 \\
0.0007875 & -0.0008942 & 0.0009948
\end{array}
$$

for one of the samples (similar results for the other).
We next evaluate the $E$ matrices of eq.(8.7) by taking discrete derivatives (taking fits with a step of 0.01 above and below the central value, making sure the fit variation is in the linear regime, taking the difference and dividing by the step size). We then substitute the $E$ and $e$ matrices into eq.(8.7) to obtain:

Corrected Error Matrix:

$$
\begin{array}{rrr}
0.00338465 & -0.00200101 & 0.00111304 \\
-0.00200101 & 0.00205966 & -0.00120513 \\
0.00111304 & -0.00120513 & 0.00114036
\end{array}
$$

Entry by Entry Ratio of Corrected to Original Minuit Error Matrix:

| 1.42272 | 1.4638 | 1.41338 |
| :--- | :--- | :--- |
| 1.4638 | 1.38418 | 1.34772 |
| 1.41338 | 1.34772 | 1.14632 |

which shows an increase in the error of about $20 \%$ for each of the diagonal entries in the end (once the square root is taken).

As discussed in Section 6.6, the bias map blows up the statistical error component of the total error, not the systematic error. Thus, the corrected error matrix is actually the one we use for our ultimate statistical error, not the original Minuit error matrix.

### 8.2 Intrinsic Errors to Efficiency Moment Method

### 8.2.1 From variation in parameters for calibration file

Because of our method of acceptance correction, there should be no dependence of the calibration function on the R parameters in the file used to derive the $\xi^{\prime} \mathrm{s}$, and in
this section we check that assumption. This must be done with samples created with various sets of parameters.

We may check the results from this effect of the variation of the parameters in EvtGen samples without smearing; the results are shown in Table 8.1. Specifically in this table what we are showing is the variation of each of the fit parameters on the file with each of the input R parameters set to 1.00 . As we iterate across the columns the calibration is taken in turn from each of the 5 canonical EvtGen files with all the R parameters set to: $0.50,0.75,1.00,1.25,1.50$.
(We have also checked the inverse effect that calibrating on the five samples and fitting to any given set of input parameters gives roughly similar behavior, so we only give in the table the results of fitting to one set of inputs, the $R_{1}, R_{2}, \rho^{2}=1.00$ file.)

As this table demonstrates, all of the fits fall within one or two sigma, with less than a third leaking out to two sigma, and we thus assign no error from this source.

| Parameter | Input | 0.500 | 0.750 | 1.000 | 1.250 | 1.500 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| $R_{1}$ | 1.000 | 0.994 | 0.997 | 1.000 | 0.997 | 0.988 |
| $R_{2}$ | 1.000 | 1.019 | 1.023 | 1.000 | 1.021 | 1.018 |
| $\rho^{2}$ | 1.000 | 0.984 | 0.980 | 1.000 | 0.987 | 0.994 |

Table 8.1: Varying the parameters used in calibration file and fitting to the file with all the input R parameters set to 1.000 (EvtGen). The numbers in the topmost row refer to the values of the parameters used to generate the calibration file (the same for all three parameters $R_{1}, R_{2}, \rho^{2}$, here). All files are on the order of 100 k events, no cuts, and the quoted Minuit errors are $\sim .015$ for each parameter

### 8.2.2 Efficiency Moment Statistical Error (EMSE) With Correlated Parameters

Using moments to correct for the acceptance/efficiency effects of the detector and reconstruction introduces an error from the MC sum done to evaluate the moments (which we denote the "EMSE error"). This contribution to the MC statistical error is not an inherent property of the moments method as such, which in the end is just a device to speed up the evaluation of the MC sum for the average efficiency $\bar{\epsilon}(\mu)$, but it simply arises from the statistical error on the MC sum for this average efficiency.

To derive an expression for this error, we begin with a linear expansion of the likelihood:

$$
\begin{equation*}
\ln \mathcal{L}=\ln \mathcal{L}_{\mu_{G}}+\left.\sum_{\alpha}^{3} \frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}\right|_{\mu_{G}} \Delta \mu_{\alpha}+\left.\frac{1}{2!} \sum_{\alpha, \beta}^{3,3} \frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right|_{\mu_{G}} \Delta \mu_{\alpha} \Delta \mu_{\beta} \tag{8.12}
\end{equation*}
$$

where the expansion is done about some initial guess $\mu_{G}$, and the sums here are over the parameter indices. The derivatives are evaluated at $\mu=\mu_{G}$ and $\Delta \mu_{\alpha} \equiv \mu_{\alpha}-\mu_{G, \alpha}$, which equals the difference between the expansion in the $\alpha^{t h}$ parameter and the central point about which we are expanding for that parameter.

To find the maximum likelihood, we solve the equations

$$
\begin{equation*}
0=\frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}=\left.\frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}\right|_{\mu_{G}}+\left.\sum \frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right|_{\mu_{G}} \Delta \mu_{\beta} \tag{8.13}
\end{equation*}
$$

for the deviations $\Delta \mu$ from the initial guess.
To compute contributions to the efficiency moment error we need the variations
from the ideal MC sample:

$$
\begin{equation*}
0=\left.\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}\right|_{\mu_{G}}+\left.\sum_{\beta} \delta \frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right|_{\mu_{G}} \Delta \mu_{\beta}+\left.\sum_{\beta} \frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right|_{\mu_{G}} \delta\left(\Delta \mu_{\beta}\right) \tag{8.14}
\end{equation*}
$$

where $\delta\left(\Delta \mu_{\beta}\right)$ is the amount we need to change $\Delta \mu_{\beta}$ by to preserve the maximum when the likelihood derivatives have varied by an amount $\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}$. If $\mu_{G}$ is well-chosen (so that the $\Delta \mu^{\prime}$ s are small) the second term in eq.(8.14) is second order and can be neglected.

Solving for the variations in the parameters $(\delta \mu)$ in this approximation, we find:

$$
\begin{equation*}
\delta\left(\Delta \mu_{\beta}\right)=-\left.\sum_{\alpha} \delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}\right|_{\mu_{G}} \times\left[\left.\frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right|_{\mu_{G}}\right]^{-1} \tag{8.15}
\end{equation*}
$$

Using $\delta \mu_{\beta}=-\delta\left(\Delta \mu_{\beta}\right)$ (since we choose $\mu_{G}$ and thus it is a constant and $\delta \mu_{G}=$ 0 ), we then find:

$$
\begin{equation*}
\delta \mu_{\beta}=-\delta\left(\Delta \mu_{\beta}\right)=\sum_{\alpha} \mathbf{e}_{\alpha \beta} \delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}} \tag{8.16}
\end{equation*}
$$

where the derivatives are evaluated at $\mu_{G}$, but we drop the notation $\left.\right|_{\mu_{G}}$ to indicate this henceforth, and generally assume it unless stated otherwise. Further, we have introduced here the error matrix Minuit will report, $\mathbf{e}_{\alpha \beta}^{M}$, whose inverse is given by $\left(\mathbf{e}_{\alpha \beta}^{M}\right)^{-1}=$ $\left(\frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right)$, or equivalently, $\mathbf{e}_{\alpha \beta}^{M}=\left(\frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right)^{-1}$.

Taking the ensemble average over a set of MC experiments of the product of two independent variations of the parameters, we then find the contribution of the MC statistics used to evaluate the efficiency is given by:

$$
\begin{equation*}
\mathbf{e}_{\alpha \alpha^{\prime}}^{E M S E}=\left\langle\delta \mu_{\alpha} \delta \mu_{\alpha^{\prime}}\right\rangle=\sum_{\beta \beta^{\prime}} \mathbf{e}_{\alpha \beta}^{M} \mathbf{e}_{\alpha^{\prime} \beta^{\prime}}^{M}\left\langle\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}} \times \delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta^{\prime}}}\right\rangle_{\text {MCsets }} \tag{8.17}
\end{equation*}
$$

where the angular brackets indicate that we take the average over the infinite set of possible MC samples that could be used to evaluate the efficiency.

### 8.2.2.1 Numerically Evaluable Form

We now substitute into this the actual likelihood in the form we are using, to obtain an expression we can evaluate numerically.

Dropping the efficiency in the numerator which plays no part in the evaluation of the best-fit parameters (as is the case in Section 6.2.1, see discussion there), the likelihood we use is given by (for notational simplicity, we omit the parameter index $\alpha$ for the next few equations, until relevant again):

$$
\begin{equation*}
\ln \mathcal{L}(\mu)=\sum_{\text {DATA }} \ln \frac{g^{t h}\left(y_{i} ; \mu\right)}{\bar{\epsilon}(\mu)}=\sum \ln g\left(y_{i} ; \mu\right)-N_{\text {DATA }} \ln \bar{\epsilon}(\mu) \tag{8.18}
\end{equation*}
$$

where $N_{\text {DATA }}$ is the number of data points, $y_{i}$ are the measured kinematic variables at data point $i$ and $g^{t h}$ is the PDF without taking efficiency into account. Variations over the first term will yield back the standard (Minuit-quoted) error, so only the variations due to the $\left[-N_{\text {DATA }} \ln \bar{\epsilon}(\mu)\right]$ term matter to us for the extra EMSE error henceforth:

$$
\begin{equation*}
\ln \mathcal{L}(\mu) \rightarrow \sum_{\text {DATA }}\left[-N_{\mathrm{DATA}} \ln \bar{\epsilon}(\mu)\right] \tag{8.19}
\end{equation*}
$$

Further, to simplify the calculation in the end, we can expand this by substituting in the explicit form of $\bar{\epsilon}(\mu)$, which we obtain from a Monte Carlo summation as follows (see eq.(6.21) ):

$$
\begin{equation*}
\bar{\epsilon}(\mu)=\frac{1}{N_{\mathrm{gen}}} \sum_{k}^{N_{\text {obs }}} \frac{g\left(\vec{y}_{j} ; \vec{\mu}\right)}{g\left(\vec{y}_{j} ; \vec{\mu}_{M C}\right)}=\frac{1}{N_{\mathrm{gen}}} \sum_{k}^{N_{\text {obs }}} w_{k}(\mu)=\frac{1}{N_{\mathrm{gen}}} W(\mu) \tag{8.20}
\end{equation*}
$$

where we have defined:

$$
\begin{equation*}
w_{k}(\mu) \equiv \frac{g\left(\vec{y}_{j} ; \vec{\mu}\right)}{g\left(\vec{y}_{j} ; \vec{\mu}_{M C}\right)} \text { and } W(\mu) \equiv \sum_{k}^{N_{\text {obs }}} w_{k}(\mu) \tag{8.21}
\end{equation*}
$$

$N_{\text {gen }}$ is the number of MC events generated with parameters $\mu_{M C}$, and $N_{\text {obs }}$ is the number of those generated MC events which actually are observed (i.e., pass all acceptance and efficiency cuts etc.). The index $k$ runs over all accepted MC events. Using this, eq.(8.19) becomes:

$$
\begin{equation*}
\ln \mathcal{L}(\mu)=-N_{\text {DATA }}\left[\ln W(\mu)-\ln N_{\text {gen }}\right] \tag{8.22}
\end{equation*}
$$

Since the second term in has no dependence on the parameters, we may drop it henceforth in finding the minimum of $\vec{\mu}$.

The variation in the first derivative of this expression is then given by:

$$
\begin{equation*}
\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\alpha}}=-N_{\mathrm{DATA}} \delta \frac{\partial \ln W\left(\mu_{\alpha}\right)}{\partial \mu_{\alpha}}=-N_{\mathrm{DATA}} \delta\left[\frac{1}{W\left(\mu_{\alpha}\right)} \frac{\partial W\left(\mu_{\alpha}\right)}{\partial \mu_{\alpha}}\right] \tag{8.23}
\end{equation*}
$$

which, dropping henceforth the assumed argument $\mu$ of $W(\mu)$ for simplicity, and expanding the $\delta$ variation, becomes:

$$
\begin{equation*}
N_{\mathrm{DATA}} \times \frac{1}{W}\left(\frac{1}{W} \frac{\partial W}{\partial \mu_{\alpha}} \delta W-\delta \frac{\partial W}{\partial \mu_{\alpha}}\right) \tag{8.24}
\end{equation*}
$$

From eq.(8.24) we can evaluate the average needed for eq.(8.17):

$$
\begin{gather*}
\left\langle\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}} \times \delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta^{\prime}}}\right\rangle_{\text {MCsets }}=  \tag{8.25}\\
\frac{N_{\mathrm{DATA}}{ }^{2}}{W^{2}} \times\left\{\frac{1}{W^{2}} \frac{\partial W}{\partial \mu_{\beta}} \frac{\partial W}{\partial \mu_{\beta^{\prime}}}\left\langle(\delta W)^{2}\right\rangle+\left\langle\delta \frac{\partial W}{\partial \mu_{\beta}} \times \delta \frac{\partial W}{\partial \mu_{\beta^{\prime}}}\right\rangle\right. \\
\left.-\frac{1}{W} \frac{\partial W}{\partial \mu_{\beta}}\left\langle\delta W \times \delta \frac{\partial W}{\partial \mu_{\beta^{\prime}}}\right\rangle-\frac{1}{W} \frac{\partial W}{\partial \mu_{\beta^{\prime}}}\left\langle\delta W \times \delta \frac{\partial W}{\partial \mu_{\beta}}\right\rangle\right\}
\end{gather*}
$$

Now we need to evaluate the averages that enter this sum of terms.
Going back to the definition of $W$ in eq.(8.21), we find (all sums here are over the MC events):

$$
\begin{gather*}
\left\langle(\delta W)^{2}\right\rangle=\sum_{j}^{N_{o b s}} w_{j}^{2}  \tag{8.28}\\
\left\langle\delta W \times \delta \frac{\partial W}{\partial \mu_{\beta}}\right\rangle=\sum_{j}^{N_{o b s}} w_{j}\left(\frac{\partial w_{j}}{\partial \mu_{\beta}}\right)  \tag{8.29}\\
\left\langle\delta \frac{\partial W}{\partial \mu_{\beta}} \times \delta \frac{\partial W}{\partial \mu_{\beta^{\prime}}}\right\rangle=\sum_{j}^{N_{o b s}}\left(\frac{\partial w_{j}}{\partial \mu_{\beta}}\right)\left(\frac{\partial w_{j}}{\partial \mu_{\beta^{\prime}}}\right) \tag{8.30}
\end{gather*}
$$

Equations (8.28) to (8.30) provide us with all the pieces we need to evaluate the contribution of the Monte Carlo integration computation of $W$ to the error.

Technically, we evaluate these in a loop over the MC events used for the $W$ calculation, the results are substituted into eq. (8.25) which is in turn substituted into eq. (8.17) to obtain the needed error matrix.

After evaluating all factors and terms, eq.(8.17) gives the ultimate full covariance matrix for the three parameters in this analysis, and the square root of the diagonal terms gives the EMSE contribution of the error to each parameter separately:

$$
\begin{gather*}
\sigma_{R_{1}}^{\mathrm{EMSE}}=\sqrt{<\delta \mu_{1} \delta \mu_{1}>} \\
\sigma_{R_{2}}^{\mathrm{EMSE}}=\sqrt{<\delta \mu_{2} \delta \mu_{2}>}  \tag{8.31}\\
\sigma_{\rho^{2}}^{\mathrm{EMSE}}=\sqrt{<\delta \mu_{3} \delta \mu_{3}>}
\end{gather*}
$$

### 8.2.3 EMSE error - Full MC Results

As this is a fairly involved computation, it was desirable to initially check the results on EvtGen samples where we were isolate this effect. A several step generatorlevel validation procedure was implemented first on EvtGen and is covered in detail in Appendix L (specifically, L.1).

Since we cannot easily run a ToyMC study on SPx MC ( $x=3$ or 4), as we cannot generate multiple samples with large numbers of events, we assume that having seen the validation of the correspondence between ToyMC and analytic calculation in Section L.1.2, for the full realistic MC we may take the analytic calculation results as a correct estimator of the EMSE error for these samples.

We now will first work with the same 9 k accepted signal MC event sample from SP3 for both the fit and calibration files - and note that the size of the fit file is not significant, because its effect is factored out into $N_{\text {DATA }}$, which cancels with the error matrix -
as $N_{\text {DATA }}$ grows, the individual terms in the error matrix grow smaller proportionately, which is essentially the central limit theorem statistical $1 / \sqrt{N}$ effect (the error from the data file size goes into the standard Minuit error, which we have dropped in eq.(8.19) to focus on the EMSE error here, which can depend only on the calibration file, and thus on the number of events $N_{\text {obs }}$ that are accepted in the generated MC file).

We then run the fit, extract the error matrix, and then run the full machinery from the last section of the analytic EMSE error estimation method on the calibration file, and obtain a diagonal EMSE error estimate for this calibration file of: ( $0.049,0.042,0.034$ ), which is not so far off from 3 times the errors on the 111 k size EvtGen file. This is roughly what we would expect from simple scaling (i.e., $\sqrt{111 / 9} \sim 3$ ), which bolsters our sense of confidence in the accuracy and applicability of this method of EMSE error estimation to both to EvtGen and full MC samples.

We go next to a SP4 signal MC calibration file, of size 15 k events, and fit to a different SP4 signal MC file (of similar size), extract the error matrix, and again use the analytic EMSE error estimation method of the last section, to obtain a diagonal EMSE error estimate of: $(0.038,0.033,0.029)$. Note that this is also scaled down from the SP3 errors by roughly $\sqrt{\frac{9}{15}}$ as we would a priori expect from statistical arguments.

Finally, we go to a larger SP4 calibration file, of size 45k events, and fit to the same SP4 sample as above, extract the error matrix, and use the same analytic EMSE error estimation method to obtain a diagonal EMSE error estimate of: ( $0.022,0.019,0.17$ ), and, again, this is also scaled down from the above SP4 fit by roughly $\sqrt{\frac{15}{45}}$, as we expect.

Going to our full calibration file of 2000-2002 generic MC, we find a diagonal EMSE error contribution to the MC statistical error of size: $(0.019,0.018,0.017)$.

### 8.3 Intrinsic DUBS Method Errors

There is also a contribution to the MC statistical error due to the use of our Direct Unbinned Background Subtraction method (DUBS) for which we derive the multiple parameter correlated expression in this section.

Take our initial pseudo-likelihood - which again is not a real likelihood because we are taking a difference of two likelihood quantities here, which is not allowed by standard maximum likelihood procedures - then from eq.(6.30) we find:

$$
\begin{equation*}
\ln \mathcal{L}(\mu)=\sum_{D A T A} \ln g_{i}(\mu)-\eta \sum_{B K G D} \ln g_{i}(\mu) \tag{8.32}
\end{equation*}
$$

where $g_{i}(\mu)$ is the value of (he (observed) PDF for event $i$ when the parameters are $\mu$, and $\eta$ is the luminosity ratio of the Data to the Background control sample which we subtract. That is, if we have $B$ background events in the control sample, then we expect $\eta B$ background events in our Data sample. After all cuts, with $N$ reconstructed Data events in the signal region, the number of expected signal events is then $S=N-\eta B$.

Since the pseudo-likelihood above is not a proper likelihood, used in a fit, Minuit will return errors as if there were $S$ signal events without properly accounting for the fact that there are $\approx \eta B$ background events. It will also do the fit without accounting
for the fluctuations in the background subtraction term.
Physically, the DUBS error piece then corresponds to a statistical variation of two parts: that in the background piece of the data (which amounts to about $20 \%$ of the total number of events) and that in the subtracted background control sample.

We begin the error derivation by reviewing in brief how we obtain the best estimate for a single parameter $\mu$ : as in Section 8.2.2, we expand to first order around some initial guess $\mu=\mu_{G}$, and solve for the value of $\Delta \mu=\mu-\mu_{G}$ that maximizes the pseudo-likelihood $(\ln \mathcal{L}(\mu))$. We then find, as in eq.(8.16):

$$
\begin{equation*}
\mu_{\alpha}=-\delta\left(\Delta \mu_{\alpha}\right)=\sum_{\beta} \mathbf{e}_{\alpha \beta}^{M} \delta\left(\frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}}\right) \tag{8.33}
\end{equation*}
$$

As before for the EMSE error, $\mathbf{e}_{\alpha \beta}^{M}$, the statistical error matrix which Minuit will report, is given by $\left(\mathbf{e}_{\alpha \beta}^{M}\right)^{-1}=\left(\frac{\partial^{2} \ln \mathcal{L}}{\partial \mu_{\alpha} \partial \mu_{\beta}}\right)$.

Taking now the variation $\delta$ of the fit parameters with respect to multiple MC trials, we find:

$$
\begin{equation*}
\delta \mu_{\alpha}=\delta\left(\Delta \mu_{\alpha}\right)=\sum \mathbf{e}_{\alpha \beta}^{M} \delta\left(\frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}}\right) \tag{8.34}
\end{equation*}
$$

The final error matrix we want to find is given by:

$$
\begin{equation*}
\mathbf{e}_{\alpha \alpha^{\prime}}^{D U B S} \equiv\left\langle\delta \mu_{\alpha} \delta \mu_{\alpha^{\prime}}\right\rangle=\sum \mathbf{e}_{\alpha \beta}^{M} \mathbf{e}_{\alpha^{\prime} \beta^{\prime}}^{M}\left\langle\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}} \times \delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta^{\prime}}}\right\rangle \tag{8.35}
\end{equation*}
$$

It is convenient for the purposes of deriving formulas for the averages needed
by eq.(8.35) to rewrite eq.(8.32) as a sum over bins:

$$
\begin{equation*}
\ln \mathcal{L}=\sum_{\text {bins }}^{\sum}\left(n_{i}-\eta b_{i}\right) \ln g_{i}(\mu) \tag{8.36}
\end{equation*}
$$

where $g_{i}(\mu) \equiv g\left(\bar{y}_{i} ; \mu\right)$ ( $\bar{y}_{i}$ is average of the kinematic variables in the $i^{\prime}$ th bin), $n_{i}$ is number of data points in the bin and $b_{i}$ the number of background control events in the bin.

The derivatives are then given by

$$
\begin{equation*}
\frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}}=\sum\left(n_{i}-\eta b_{i}\right) \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta}} \tag{8.37}
\end{equation*}
$$

The variation is then

$$
\begin{equation*}
\delta\left(\frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}}\right)=\sum\left(\delta n_{i}-\eta \delta b_{i}\right) \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta}} \tag{8.38}
\end{equation*}
$$

We now use $<\delta n_{i} \delta b_{i}>=0,<\delta n_{i} \delta n_{j}>=n_{i} \delta_{i j}$ and $\left.<\delta b_{i} \delta b_{j}\right\rangle=b_{i} \delta_{i j}$ (Poisson statistics) to evaluate the MC ensemble averages - i.e., there is no correlation between the amount of variation of events in any two bins, and the variation in a given bin is given by $\sqrt{n}$. We use these to evaluate the log-squared average piece of eq.(8.35):

$$
\begin{equation*}
\left\langle\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}} \times \delta \frac{\delta \partial \ln \mathcal{L}}{\delta \partial \mu_{\beta^{\prime}}}\right\rangle=\sum\left(n_{i}+\eta^{2} b_{i}\right) \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta}} \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta^{\prime}}} \tag{8.39}
\end{equation*}
$$

This can be rewritten as:

$$
\begin{equation*}
\left\langle\delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta}} \times \delta \frac{\partial \ln \mathcal{L}}{\partial \mu_{\beta^{\prime}}}\right\rangle=\sum\left(\left(n_{i}-\eta b_{i}\right)+\eta(1+\eta) b_{i}\right) \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta}} \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta^{\prime}}} \tag{8.40}
\end{equation*}
$$

where the first term is just usual contribution to the error that is provided by Minuit, and the second term the extra error contributed by the DUBS procedure.

The second term includes the contribution due the presence of background in the Data sample (which comes in with coefficient $\eta$ ) and the contribution of the variation of the background in the background control sample which is subtracted (this comes in with coefficient $\eta^{2}$ ). To evaluate it we substitute the second term into eq.(8.35) to obtain

$$
\begin{equation*}
\mathbf{e}_{\alpha \alpha^{\prime}}^{D U B S}=\eta(1+\eta) \sum_{\beta \beta^{\prime}}^{\mathbf{e}_{\alpha \beta}^{M} \mathbf{e}_{\alpha^{\prime} \beta^{\prime}}^{M}} \sum_{b k g d} \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta}} \frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta^{\prime}}} \tag{8.41}
\end{equation*}
$$

where we have now converted back to a sum over events (in direct analogy to the transition from eq.(6.15) to eq.(6.16), where the occupancy function becomes binary so that the transition from a sum over bins to one over events results in simply a sum over the accepted events, we here are absorbing the $b_{i}$ of the second term of eq.(8.39) as a binary function so that eq.(8.41) becomes a sum over the accepted background events).

This formula shows the expected behavior. If there is no background then we sum over zero events and there is no extra contribution to the error. If the MC background control sample is very large, i.e., $(\eta \ll 1)$, the contribution from it becomes very small. However, it does not go to zero, because in fact $\eta \ll 1$ implies that the last factor is a sum over the background events, (i.e., $\Sigma b_{i}$ ), which is becoming larger and larger as $\eta$ grows, and in fact the product of $\eta$ and this factor will be relatively constant however many events are being subtracted. What this corresponds to is the residual error from statistical variation in the number of background events in the data sample itself, which cannot be eliminated by subtracting larger sets of MC, but only by increasing the amount of events in the data sample itself (this would be manifested in practice in the terms in
the error matrix becoming smaller as $N_{\text {DATA }}$ grows, so that the denominator of eq.(8.41) then grows, making the entire expression smaller).

Thus we see that these two errors of statistical variation in the number of background events in the data sample and in the subtracted MC control sample enter eq.(8.41) in the same form, just with different coefficents: $\eta$ for the former, and $\eta^{2}$ for the latter. We will later in fact split these errors up, into an actual irreducible data statistical error for the former (which we will refer to as the DUBS irreducible error) that cannot be affected by the size of the MC sample, and a MC statistical error for the latter that in theory can be reduced to zero for a large enough subtracted MC control sample.

For $\eta=1$ as we will have below in our EvtGen validation, the size of both are equivalent.

Finally, we have been implicitly using all along the notation $g_{i}=g_{i}^{\text {obs }}$, thus from eq.(6.9), $g_{i}(\mu)=g_{i}^{t h} \epsilon(\vec{y}) / \bar{\epsilon}(\mu)$ where from eq.(8.20), $\bar{\epsilon}(\mu)=W(\mu) / N_{\text {gen }}$. So, using again the fact that neither $\epsilon(\vec{y})$ nor $N_{\text {gen }}$ depend on the parameters $\mu_{\alpha}$, we may expand the derivatives in eq.(8.41):

$$
\begin{equation*}
\frac{\partial \ln g_{i}(\mu)}{\partial \mu_{\beta}}=\frac{1}{g_{i}^{t h}(\mu)} \frac{\partial g_{i}^{t h}(\mu)}{\partial \mu_{\beta}}-\frac{1}{W} \frac{\partial W}{\partial \mu_{\beta}} \tag{8.42}
\end{equation*}
$$

where from eq.(8.21), $W=\sum_{k}^{N_{o b s}} w_{k}(\mu)$. This is the final form we evaluate for the derivatives over the background events to give us the error estimate.

Note that neither the EMSE nor DUBS errors can be written as simply a factor times the Minuit error because the contribution include correlations that which in general
will be different than those present in a background-free or perfect efficency fit.
Also though it may initially appear curious that to obtain the error from the background we must evaluate background events over the signal PDF, on closer examination, this is in fact not so unexpected as the entire DUBS method relies on subtracting the contribution of the background events to the likelihood by subtracting their individual contributions one by one from the total likelihood as evaluated over the data events, so it can be expected to follow intuitively that the statistical error due to this background subtraction will eventually be manifested in a formula which has pieces that need to be evaluated over signal PDF itself.

Further, this DUBS error takes into account all errors associated with the statistical rate variation in both the combinatorial and peaking backgrounds, so that these errors do not have to be computed separately.

After evaluating all factors and terms, eq.(8.41) gives the ultimate full covariance matrix for the three parameters in this analysis from the DUBS error source, and the square root of the diagonal terms gives the DUBS contribution of the error to each parameter separately:

$$
\begin{gather*}
\sigma_{R_{1}}^{\text {DUBS }}=\sqrt{<\delta \mu_{1} \delta \mu_{1}>} \\
\sigma_{R_{2}}^{\text {DUBS }}=\sqrt{<\delta \mu_{2} \delta \mu_{2}>}  \tag{8.43}\\
\sigma_{\rho^{2}}^{\text {DUBS }}=\sqrt{<\delta \mu_{3} \delta \mu_{3}>}
\end{gather*}
$$

### 8.3.1 DUBS error - Analytic Results on Generator-level MC Samples

As with the EMSE validation, it is desirable to check these analytic formulas results on EvtGen samples where we can isolate this effect, and careful studies were done on this and are shown in detail in Appendix L.2.

### 8.3.2 DUBS error - Full MC Results

As with EMSE, since we cannot run a toyMC study on SPx MC (x $=3$ or 4$)$, as we cannot generate many multiple samples with large numbers of events, we assume that having seen the validation of the correspondence between toyMC and analytic calculation in the last section, for the full realistic MC we may take the analytic calculation results as a correct estimator of the DUBS error for these samples.

We now take the standard full 11k accepted MC event sample after all cuts from SP3. This consists of about 9 k MC truth selected signal, and 2 k MC truth selected background. We use the former as the calibration file, and we do the standard subtraction of the latter from the full sample, leaving us with the 9 k signal sample which we then fit.

We then run the fit for the signal sample as above and extract the error matrix, and then take the background sample of about $2 k$ events, run the full machinery from the last section of the analytic DUBS error estimation method on it, and obtain a diagonal DUBS error estimate for this file of: $(0.056,0.040,0.030)$, which is about a factor of two larger than 3 times the errors on the 20k size EvtGen file, 3 being the factor we might expect from simple statistical scaling (i.e., $\sqrt{20 / 2} \sim 3$ ). However, in this case there is no
reason to expect that the DUBS error will scale simply numerically, since the distributions of the backgrounds will in general be significantly different in space from that of the simplistic single "background parameter" set we generated in the EvtGen tests (see discussion at the beginning of Section L.2). We thus have no a priori sense of how the much the different types of background will contribute to this error.

We might however expect that the SP4 MC distributions are not so radically different from those of SP3, and we should see some level of scaling there, and indeed when we use the same procedure to run the analytic calculation on the standard test SP4 sample of total 19.5 k events, 15.5 k signal, 4 k background, we find a diagonal DUBS error estimate of: $(0.042,0.030,0.025)$, which indeed is roughly a factor of $\sqrt{\frac{2 k}{4 k}}=0.71$ down from the SP3 errors.

Finally, we go to a larger SP4 calibration file, of signal size 45 K events, with 6 k background events, and fit to the same SP4 sample as above, extract the error matrix, and use the same analytic DUBS error estimation method to obtain a diagonal DUBS error estimate of: $\sigma_{D U B S}=(0.022,0.019,0.017)$. Again, this is also scaled down from the above SP4 fit by roughly $\sqrt{\frac{15}{45}}$, as we expect.

This MC sample corresponds to a data-luminosity equivalent of $51 \mathrm{fb}^{-1}$ and for this section we will assume a dataset luminosity of size $33 \mathrm{fb}^{-1}$ (this is the initial size we did the study with, before expanding to the full $78 \mathrm{fb}^{-1} 2000-2002$ dataset). This yields a factor of $\eta=33 / 51$, and thus the size of the irreducible DUBS systematic is about $\eta * \sigma_{D U B S}=(.014, .012, .011)$ and the reducible one due to limited MC statistics is
$\eta^{2} * \sigma_{D U B S}=(.009, .008, .007)$.
At first glance the behavior of this formula with fixed MC samples size but increasing $\eta$ - that is, increasing data sample size alone - seems curious, because both pieces of the DUBS error would seem to increase, since they get $\eta$ or $\eta^{2}$ factors. However, looking more closely, we find in eq.(8.41) that the full DUBS error matrix is directly proportional to the Minuit statistical error matrix on the data alone, which of course will scale down statistically with data. Thus, the DUBS error will scale down with increasing data samples, though this is not so immediately obvious from the above paragraph alone.

Going to our full calibration file of 2000-2002 generic MC, we find a diagonal reducible DUBS error contribution to the MC statistical error of size: $(0.017,0.010,0.009)$, and a diagonal irreducible DUBS contribution to the data statistical error of size (0.031, $0.018,0.016)$, which is what is added to the Bias Map Correctioned-data statistical error of Section

### 8.4 Detector Related Errors

### 8.4.1 Tracking

Because our measurement is a relative one and does not depend on the absolute difference between yields in MC and data, and because the MC has been tuned over time to match the Data for tracking, we expected and found that this was a relatively small contribution to our error budget.

### 8.4.1.1 Lepton Tracking

We obtain a rough estimate of this contribution if we apply the correction for GTL (GoodTracksLoose tracks, from which all the leptonic and hadronic candidates other than the slow pion are chosen, see Section ??), suggested in Section 2.3 of the report by the Tracking Efficiency Task Force ([34]) to be applied as an overall correction vs. a binned one. This correction to the MC relative to data is $0.8 \%$, averaged over $p_{T}, \theta, \phi$, multiplicity, and charge. A more accurate weighted average can be derived from the full machinery of the tracking correction tables, averaging just the correction on leptons over accepted events, this was done in Ref. [42], and the averaged correction for leptons found to be $1.1 \%$ (we verified the magnitude of this number by doing averages over the tracking correction tables ourselves as well).

Since we need to make a correction that weights different bins differently, we chose to multiply the lepton spectrum (of the fit file, for an MC calibration to MC fit file) by a sloped line, under which the area for accepted leptons (from $1.2 \mathrm{GeV} /$ c to $2.4 \mathrm{GeV} / c$ ) was less or more than that for a flat line (i.e., just applying the lepton spectrum by unity, leaving an unchanged spectrum) by $1.1 \%$ (this correction factor is taken from Table 3.6 of [42]). We did this for a downward sloping line and obtained a deviation from the original fit of ( $-0.007 /-0.006 /+0.002$ ) and then for an upward sloping line, with a deviation from the original fit of $(+0.007 /+0.006 /-0.002)$. In fact, a glance at the $p_{t}$ diagrams in Figs. 4-6 of [34] shows that above $300 \mathrm{MeV} / \mathrm{c}$, the ratio of number of accepted events between Data and MC is essentially flat, so we regard the above upper limit as rather too
generous, and can with strong assuredness take the conservative step of assigning zero for all leptonic tracking corrections.

### 8.4.1.2 Hadronic Daughter Tracking

For the $D^{0}$ daughter tracking error, we did a similar study, taking the $K \pi$ mode and weighting each of the momentum spectra for the K and $\pi$ up and down in their accepted regions ( 0.1 to $3.5 \mathrm{GeV} / c$ ) by lines with $1 \%$ area differences as above for the lepton. The $1 \%$ number is taken from averaging just the correction on pions or kaons over accepted events, which was done in Ref. [42], and results in the same 1\% correction (from Table 3.6 of [42]).

As might be expected, the largest correction was obtained for the case where both $K$ and $\pi$ are weighted with lines sloping in the same direction, and this deviation (for upsloping lines) is ( $-0.0003,+0.0004,-0.0020$ ). However, note that the large bulk of these hadronic daughters occur with momentum greater than $300 \mathrm{MeV} / c$, where the Data/MC ratio is basically flat as mentioned above, and further that the decay distribution in the $D^{0}$ rest frame is flat (since the the $D^{0}$ is spinless), indicating that for the two-body decay with equal momenta magnitude in the $D^{0} \mathrm{CM}$ frame, its daughters would end up being in lower and higher momentum bins when boosted into the lab frame, so that the corrections would tend to cancel. This can be extended to the cases where more daughters are emitted, with momenta throughout the accepted range, with further cancellations in any corrections.

All this would indicate we can safely assume that the full correction is less than that of even the lepton, and we do not then assign any error from this source either.

### 8.4.2 Slow Pion Tracking

The error due to slow pion tracking is evaluated specifically differently from other tracking errors because it is not covered by the control samples used for higher momentum tracks. The slow pion from the decay $D^{*+} \rightarrow D^{*} \pi^{+}$is the most difficult of the decay products to reconstruct, due to its very low momentum (between 50 and $\sim 250 \mathrm{MeV} / \mathrm{c}$ ). It often (about $2 / 3$ of the time) fails to penetrate the drift chamber and in these cases, we are dependent entirely on the vertex detector for detection and measurement. Figure 8.1 illustrates that we do a good job of modeling the reconstruction of these difficult tracks, but we need to make a quantitative estimate of the possible error, since understanding and modeling the acceptance is critical in knowing the dependence of our efficiency on $w$.

As in the BaBar $V_{c b}$ analysis [39], we use the Menary function to model shape the slow pion efficiency as a function of momentum. We extract the parameters of this function from a fit to inclusive $D^{*}$ helicity angle distributions in both the MC and the data. We apply the ratio of these function as a weight and observe how much our answer shifts: ( $0.0025,0.0002$ and 0.0110 ) for $R_{1}, R_{2}$ and $\rho^{2}$, respectively. As might be expected, $\rho^{2}$, which is the parameter most sensitive to the shape of the $w$-distribution, is most strongly affected.


Figure 8.1: Plots for $\cos \theta_{\pi}$ (angle between the slow pion and $D^{*}$ direction) for data (points) and MC (histogram) in bins of $D^{*}$ center-of-mass momentum. In units of $\mathrm{GeV} / c$ the bins are (1) 0.5-1.0, (2) 1.0-1.5, (3) 1.5-2.0, (4) 2.0-2.5, (5) 2.5-3.0, (6) 3.0-3.5 and (7) 3.54.0. $B^{\prime}$ 's can only produce $D^{*}$ momenta has high as bin 4 . Bins 1-3 are most relevant to this analysis.

We do not apply this weight as a correction, as its origin in a fit to the inclusive $D^{*}$ distribution means it is not strictly applicable to our channel. However, it is a measure of how much the MC and data differ, so it a reasonable way to gauge our systematic. We take this shift as our systematic error from this source.

### 8.4.3 PID

This is the error due to the difference in PID efficiency in MC and data.
The PID AWG has established the misID rates using control samples, and published their PID tables for corrections to central values of results sensitive to absolute rates at [35]. Again, as we only care about relative weights, we may try to obtain an overestimate on the error by simply taking averages in these tables over all bins except for momentum (actually, the the only other binning which matters is over theta, and the tables only take one phi bin), taking the ratio per momentum bin for data over MC misID rates, then taking reasonable variations across the particle momenta spectra derived from these values, weighting the fit sample by these factors, and re-evaluating the fits. The discrepancies from the unweighted fits give us an estimate then of the systematic error due to misID data/MC discrepancy.

### 8.4.3.1 Leptonic PID

We estimated this using the table for the correction of the leptons (specifically electrons), which resulted in the values for the correction factor varying over the range of 0.991 to 1.008 over the accepted lepton momentum spectrum. We then took reasonable
maximally sloped lines within this range, reweighted the lepton spectrum for accepted events and then reran the fits: the deviations from the original came to $(+.0064,+.0052$, $-.0016)$ for an upward-sloped line and (-.0032, -.0031, +.0009) for a downward-sloped one.

### 8.4.3.2 Hadronic PID

Since there is no $\pi$ ID used in our analysis, we did not need to worry about misID error for $\pi^{\prime} \mathrm{s}$, but we do use NotAPion Kaon ID in the $D^{0} \rightarrow K \pi$ mode, and we performed a similar analysis in this case with the values for the correction factor varying over the range from 0.985 to 1.011 over the accepted Kaon momentum spectrum. We then took reasonable maximally sloped lines within this range, reweighted the Kaon spectrum for accepted events and then reran the fits: the deviations from the original came to ( -.00002 , $-.0003,+.0038)$ for an upward-sloped line and $(+.0001,+.0001,-.0037)$ for a downwardsloped one. We also did a bin-by-bin estimate in this case using the weighting factor for the Kaon momentum per bin, and found the deviations from the original came to: (+.0010, -.0002, -.0009), so slightly larger in $R_{1}$, but otherwise as expected less than for the sloped lines overestimates.

### 8.4.3.3 Total PID error

We can see from the non-monotonicity of the correction curve that both of these are an overestimate of the actual error, but adding in the hadronic daughter PID misID rate, we may safely assign an error of both lepton and Kaon PID misID rate added in quadrature to be $(0.006,0.004,0.004)$.

### 8.4.4 $\quad \delta m$ Width difference

We fit to the $\delta m$ peak in data and MC, and then attempt to adjust the width in the fit sample by the noted difference between them.

The fits of a single Gaussian plus a threshold function (see e.g. BAD 527 [42], Section 4.2.2, for details on this, including the standard form of the threshold function) to the SP3 MC and Release 8 Data $\delta m$ distribution are shown in Table 8.2, and in Fig.8.2. While the BR and $\left|V_{c b}\right|$ analyses do double Gaussian fits to the $\delta m$ peak as part of their procedures, we may fit simply to a single Gaussian for the purposes of this error (see estimate (see Appendix M for details on this).

It is clear from Table 8.2 that the differences in mean and width for the fit to $\delta m$ are at the level of tenths and hundredths of an MeV , respectively. From this we might expect that there is likely little error that arises from this source, and to check this we varied the width of our $\delta m$ from our standard value to a wider value corresponding to the addition of the difference between our fit widths of Data and MC, and then did the standard scaled subtraction of background. The results came out to be shifted with respect to the original fit by $(+0.006,-0.004,+0.008)$, which is to be compared with a statistical error for this sample of about $(0.030,0.023,0.019)$. It can be seen that the above shifts fall easily within the statistical error, and can thus be due entirely to statistical fluctuations within the newly included events - thus there is no reason to assign a separate actual systematic error from this source.

Because this is one of the errors that relies on a proper estimate of the back-
ground composition and amount, it is not easily testable in toyMC, but from the above numbers and test, we are confident that we may assign this to be a negligible source of error.

| Quantity | Mean (MeV) | Width (MeV) |
| :--- | :---: | :---: |
| Data | 145.42 | 0.589 |
| MC | 145.56 | 0.596 |

Table 8.2: Results of fits to MC and Data $\delta m$ distributions.


Figure 8.2: Distribution of $\delta m$ for MC (L) and Data (R).

### 8.4.5 Error due to FSR

This is the possible error due to the fact that one or more photons may be lost in the process of Final State Radiation (emission of a photon actually within the Feynman diagrams of the decay process itself, as opposed to later emission due to Bremstrahlung). This is not taken into account in the theoretical PDF eq.(3.5), which only assumes the final state particles $D^{*}$, lepton, and $\nu$. Thus, depending on how often it occurs and how
strongly this affects the lepton spectrum (as almost all FSR photons are emitted by the lepton, because of its lightness with respect to the other charged particles involved in the decay), it might in theory be thought to have an impact on the fit results.

We investigated this error closely in toyMC by simulating samples with FSR (i.e., PHOTOS, in EvtGen language) turned on and off, we include in Appendix N the detailed steps to our assigning a null result for the FSR error for this analysis, here we give a synopsis.

The steps we followed were these:

- 1. First took large samples with no acceptance cut and no FSR and normalized using just the analytic function, fit is fine.
- 2. We then did the same no-acceptance fit with analytic normalization to a file with FSR turned on, fit results seem off.
- 3. Went to the more realistic case of testing on files with actual acceptance requiring real calibration: began with calibration on a no-FSR sample and fitting to another no-FSR sample, fit works fine.
- 4. Then took calibration on a no-FSR sample and fitted to a sample with FSR (with similar sample sizes), ultimate deviations: $(+.0129,+.0067,-.0039)$, to be compared with statistical uncertainties on this fit on the order of (.0060, .0060, .0050) - fit results seem off, at least in $R_{1}$, though not dramatically so, and still within a couple of sigma of being correct.
- 5. Next tried calibration on samples with FSR and fitted to a sample with FSR (with similar sample sizes), ultimate deviations: (+.0003, -.0022, -.0025) (same order of statistical uncertainties as above) - ok.
- 6. Next tried calibration on a photon-resummed FSR sample and fitted to a sample with FSR (with similar sample sizes) to see if it improved the result, ultimate deviations: $(-.0053,+.0023,-.0054)$ (same order of statistical uncertainties as above) ok.

Our final conclusions: FSR does have a real effect, and either of the last two correction methods is viable, and concordant with assigning a zero error from this source.

Thus, we take the simpler technique of just doing calibration with a sample with FSR on in the MC, and will not do any resumming to obtain our final result.

To estimate the error, we take the prescription recommended by [43] where $30 \%$ of the discrepancy between the absolute values of the results with FSR on and those with FSR off (i.e., the differences between \# 4 and \# 5) are taken as the systematic errors from this source. This gives us: $(0.0126,0.0045,0.0014) * 0.3=(0.0038,0.0014,0.0004)$. Though these are small enough that we may regard them as negligible (as is generally any error that falls below a threshold of about 0.005), to be conservative, we include this in our total error budget.

### 8.4.6 Peaking Backgrounds

These are all backgrounds in which a $D^{*}$ was reconstructed correctly and thus they fall into the peaking region in the $\delta m$ plot, but other problems arose, such as the lepton not being correctly reconstructed or the original decay was a different semileptonic $B$ decay mode.

These fall into the following categories:

- Fake Lepton: the lepton that was found was not actually an $e$ or $\mu$ but a hadron (primarily $\pi^{\prime}$ s and $K^{\prime}$ s) that passed lepton ID cuts.
- Other semileptonic channels: These fall into three types - a $B^{0}$ or $B^{ \pm}$decaying to $\ell \nu$ pair along with either a $D^{0}$, a $D^{* *}$, or a nonresonant combination. Some of these events will pass all cuts and then fall into the correlated (cascade) and uncorrelated (opposite side) lepton peaking background categories as described in Section 5.2.4. For example, a charged $B$ can decay into a $\ell \nu+D^{0 * *}$ which immediately decays into a $D^{*}$ plus at least one pion, which escapes detection, so it ultimately becomes: $B^{ \pm} \rightarrow D^{*} \ell \nu+n \pi_{s o f t}$.


### 8.4.7 Other semileptonic channel contributions

These decays fall into two types, the first type are where an extra pion in the decay was missed in the reconstruction. An example of this type is a mode in which a charged or neutral $B$ meson decays to a $D^{* *} l^{ \pm} \nu$, then the $D^{* *} \rightarrow D^{* \pm} \pi$, with the charged or neutral pion not being detected, and the $D^{* \pm}$ going on to decay as normal. These type
mimic our signal mode exactly, and thus pass many of the cuts $\left(\cos \theta_{B Y}\right.$ being the one that most stringently cuts them out, since the calculated angle between the Y -system and the initial $B$ direction is not real, and will often fall outside the allowed bounds of -1 to 1 , see [36] for more detail on this). The same applies to non-resonant decays to $D^{*} \ell \nu \pi$-type final states, where the extra $\pi$ is missed.

The second type are those where an extra pion was inadvertently added into the decay chain. An example of this type is where $B \rightarrow D l^{ \pm} \nu$, where the $D$ is in the ground state. In this case, an extra pion must be accidentally added (from e.g. the other side of the decay) to the $D$ to form a particle that peaks in the $D^{*}$ region.

Both of these types of decays make it into the ultimate signal region in substantial sizes after all cuts are applied, and thus we must employ some method to determine how much error is introduced into our result from the expected uncertainty in the amounts of subtracted backgrounds of this type.

Our modeling of the distributions produced by these type of peaking backgrounds will dependent on the not very well known branching fractions of semi-leptonic $B$ decay through the mixture of modes that represent these backgrounds. To check our sensitivity to these branching fractions we vary the mixture by relatively large amounts.

To estimate the uncertainty associated with the mixture of background modes, we vary them by weighting them up and down, rerun the fit, and observe how much the parameters shift from those obtained with the nominal mixture. During this variation we keep the total fractions as determined by the $\cos \theta_{B Y}$ fits (see Section 6.4) fixed. We
vary most modes by $60 \%$. For the measured mode $D_{1} l \nu$ we vary only by the $\sim 30 \%$ measurement error[1]. In the case of $D^{*} l \nu$ self cross feed from badly reconstructed signal, we take a $20 \%$ variation of the branching fraction to account for the discrepancies in the literature. For $D^{*} l \nu$ we also use $20 \%$ - the contribution to the error being in any case very small.

Table 8.3 shows the semileptonic modes that a $B$ can decay into and the variations we assigned. Charged and neutral channels related are grouped together and were varied by a common factor ( because the underlying quark processes are related by isospin, and we assume isospin invariance is a good symmetry, which generally holds well in $B$ system decays). The effect of varying the $c \bar{c}$ contribution is also included. Each channel is varied up and down by the indicated factors, and we take half of the resulting variation in the fit parameters as our systematic error. These errors are recorded in the last column of the table.

The total error for our three parameters is ( $0.0140,0.0077$ 0.0052). It was obtained by adding the errors due to each channel in quadrature. $R_{1}$ is the most sensitive to the mixture of decay modes used in the background subtraction.

It is clear that the modes that dominate the total error are the decay to the ground-state $D^{0} / D^{ \pm}$meson, three of the $D^{* *}$ modes (modes 2-5), and the charged $B$ to $D^{*}$ mode; the nonresonant modes total contribution is relatively small.

The total error here turns out to be much smaller than in the last CLEO measurement (where it was all put into the category of " $D^{* *}$ error") which did not subtract

| Decay Mode | MC BR | PDG BR | Var. | Error from BR Variation <br> $\left(\sigma_{R_{1}}, \sigma_{R_{2}}, \sigma_{\rho^{2}}\right)$ |
| :--- | :---: | :---: | :---: | :---: |
| $0 . B \rightarrow D^{*} \ell \nu($ signal $)$ | 4.8 | $5.7 \pm 0.21$ | 20 | $(0.00052,0.00044,0.00027)$ |
| $1 . B \rightarrow D \ell \nu$ | 2.10 | $2.14 \pm 0.14$ | 20 | $(0.0013,0.00037,0.0002)$ |
| $2 . B \rightarrow D_{0} \ell \nu$ | 0.20 | - | 60 | $(0.00020,0.00026,0.00010)$ |
| $3 . B \rightarrow D_{1} \ell \nu$ | 0.56 | $0.56 \pm 0.16322$ | 30 | $(.0087, .0 .0024,0.0016)$ |
| $4 . B \rightarrow D_{1}^{\prime} \ell \nu$ | 0.37 | - | 60 | $(0.012,0.0062,0.0044)$ |
| $5 . B \rightarrow D_{2} \ell \nu$ | 0.37 | - | 60 | $(0.00095,0.0025,0.0017)$ |
| $6 . B \rightarrow D^{*} \pi \ell \nu$ | 0.30 | - | 60 | $(0.0036,0.00087,0.00071)$ |
| $7 . B \rightarrow D \pi \ell \nu$ | 0.9 | - | 60 | $(0.0022, .00092, .00061)$ |
| $20 . c \bar{c}$ | NA | NA | 20 | $(0.0011,0.00034,0.00040)$ |
| Total |  |  |  | $(0.016,0.0073,0.0051)$ |

Table 8.3: All the semileptonic $B$ decay modes. The first column is the BR (in percent) put into the MC when this mode is generated from generic $B$ decays, the PDG (2003) measurement is given if available, and the third column is the percentage variation used. The last column is the error per mode computed from the variations. Modes related by isospin between $B^{0}$ and $B^{+}$decays have been treated together.
the $D^{* *}$ but simply did a crude estimation of what effect it would have on the fit if added back in, and found it to be one of their largest errors.

### 8.4.7.1 Background $w$-dependence

The $w$-dependence of the background is taken from $\cos \theta_{B Y}$ fits in $w$-bins undertaken for the $V_{c b}$ analysis [39]. We fit this $w$ to a polynomial centered in the middle bin in order to extract this $w$-dependence for our use. We estimate the error from the $w$-dependence by varying the slope of this fit by $\pm 1 \sigma$ and repeating the likelihood fits to see how our result varies. One half the difference between these $\pm$ variations is assigned as the systematic error from this source.

### 8.4.7.2 Varying total peaking and combinatorial background fractions

The other source of background error is the overall normalization for the peaking and combinatorial fractions. To estimate these we take the measured variation in the $V_{c b}$ analysis [39], vary the $f_{\text {peaking }}$ (referred to as $\lambda_{\text {sw }}$ in Section 6.4) fraction up and down by this amount, refit in both these cases, and take the average of the difference from the original fit as an error estimate. This yields $(0.016,0.0067,0.0017)$ for the three parameters, indicating this is a significant error source for $R_{1}$ and $\rho^{2}$, especially.

Following a similar procedure with $f_{\text {comb }}$ (referred to as $\lambda_{\mathrm{sb}}$ in Section 6.4) we find an error estimate of ( $0.0027,0.0014,0.0035$ ), indicating that this is a much smaller source of error than the peaking variation, for all three parameters.

### 8.4.8 Combinatoric $\delta m$ background difference in shape

In this section we cover the second primary contribution to the MC statistical error from the DUBS method.

As we saw in Chapter 7, the agreement of the distributions of the kinematic variables between Data and MC is quite good in the sideband region (which contains mostly combinatoric background). But since the combinatorial background comprises about half of the total background under the peak, the small differences in the shapes of the distributions may yet introduce some error when the MC is used to subtract the remaining combinatorial background in the signal region from the data. To investigate this effect, we took the ratios of Data to MC in the sideband region and took the best
polynomial fits to the points, with the result shown in Figure 8.3.


Figure 8.3: Ratio of Data to MC for the four kinematic variables in the high sideband region in Release 10 data, and best polynomial fits (varying in order from 4 to 6 depending on what best fitted the distribution).

The $\frac{\chi^{2}}{d o f}$ given are for the fits with the functions. We chose a fourth order polynomial for $w$ here and a sixth order polynomial for the other kinematic variables. These gave a reasonable $\frac{\chi^{2}}{d o f}$.

We next use these functions one at a time to multiply the combinatoric background from MC before it is subtracted from the data to prepare the sample for the fit for the form factors. We compare the form factors we get from these fits with the form factors we get from the fits with the unaltered background. This procedure yields the
results shown in Table 8.4.
The differences we find are generally small (the largest is from using the function we found for $w$, from which we find $\Delta \rho^{2} \sim 0.006$ ), adding them in quadrature we assign from this source ( $.005 ; .004 ; .006$ ).

| Reweighted distributions | $R_{1}$ | $R_{2}$ | $\rho^{2}$ |
| :--- | :---: | :---: | :---: |
| w distribution | -0.002 | 0.0 | 0.006 |
| $\cos \theta_{\ell}$ distribution | 0.001 | -0.002 | -0.001 |
| $\cos \theta_{V}$ distribution | 0.002 | -0.003 | 0.001 |
| $\chi$ distribution | 0.004 | -0.002 | 0.001 |

Table 8.4: Changes in the fitted parameters for reweighting of the MC combinatoric background distributions in the four kinematic variables, as shown in Figure 8.3.

### 8.5 Final Errors

The above descriptions include all significant errors at this stage, and in Table 8.5 we compile all the systematic error estimates from this Chapter, and estimate a total estimates for both systematic errors and for MC statistical errors for our $78 \mathrm{fb}^{-1}$ of data, $K \pi e$ mode result.

Corrected (Minuit + Bias Map Correction error) Statistical for $78 \mathrm{fb}^{-1}$ data:
(0.055; $0.044 ; 0.039)$

| Error Source | Estimates for $\left(R_{1} ; R_{2} ; \rho^{2}\right)$ |
| :--- | :--- |
| Lepton Tracking Efficiency | $(0.005 ; 0.004 ; 0.002)$ |
| Slow Pion Tracking Efficiency | $(0.003 ; 0.0002 ; 0.011)$ |
| PID efficiency (lepton, kaon) | $(0.006 ; 0.004 ; 0.004)$ |
| Final state radiation | $(0.0043 ; 0.0023 ; 0.0013)$ |
| Peaking background normalization | $(0.019 ; 0.008 ; 0.001)$ |
| Combinatorial background normalization | $(0.006 ; 0.003 ; 0.008)$ |
| Background composition (branching fractions) | $(0.016 ; 0.0079 ; 0.0051)$ |
| $w$-dependence of background | $(0.001 ; 0.001 ; 0.028)$ |
| Combinatorial background shape | $(0.005 ; 0.004 ; 0.006)$ |
| Total Systematic | $(0.021 ; 0.012 ; 0.033)$ |

Table 8.5: Systematic errors estimates table (errors contributing negligible amounts: $D^{0}$ tracking and DeltaM MC/Data differences, are not shown in this table; see text) .

| MC Statistics Error Source | Estimates for $\left(R_{1} ; R_{2} ; \rho^{2}\right)$ |
| :--- | :--- |
| EMSE - Eff. Moment Statistical Error | $(0.019 ; 0.018 ; 0.017)$ |
| DUBS Contrib to MC Statistical Error | $(0.017 ; 0.010 ; 0.009)$ |
| Total MC Statistical | $(0.026 ; 0.020 ; 0.019)$ |

Table 8.6: MC statistical errors for our $\sim 200 \mathrm{fb}^{-1}$ equivalent SP4 generic MC sample.

| Data Type and Error Source | Estimates for $\left(R_{1} ; R_{2} ; \rho^{2}\right)$ |
| :--- | :--- |
| SP3 or SP4 MC Stat'l Error | $(.17 ; .15 ; .13)$ |
| EvtGen Stat'l Error | $(.11 ; .12 ; .12)$ |
| SP3 Stat'l Error | $(.16 ; .14 ; .11)$ |
| SP4 Stat'l Error | $(.17 ; .13 ; .10)$ |
| CLEO Stat'l Error | $(.29 ; .21 ; .16)$ |

Table 8.7: MC statistical and normal statistical errors per unit $\mathrm{fb}^{-1}$ of MC luminosityfor the MC statistical error, or per 1 k events in the case of the actual statistical error (luminosity cannot be directly compared between CLEO and Babar because CLEO only reconstructed two $D^{0}$ decay modes, where Babar reconstructs four, but for Babar roughly 1 k signal events are reconstructed per $\mathrm{fb}^{-1}$ ). At a naive level, statistical errors at a collected data luminosity $\mathcal{L}_{2}$ can be scaled from the above SP4 statistical errors simply by multiplying by the factor $\sqrt{\frac{1}{\mathcal{L}_{2}}}$, while MC statistical errors at a MC equivalent luminosity of $\mathcal{L}_{M C}$ can be scaled from the above SP4 MC statistical errors by multiplying by the factor $\sqrt{\frac{1}{\mathcal{L}_{M C}}}$

## Chapter 9

## Conclusions

Thus we quote finally that we have determined the three form factors involved in $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ decay to be:

$$
\begin{align*}
& R_{1}=1.34 \pm 0.055 \pm 0.026 \pm 0.025,  \tag{9.1}\\
& R_{2}=0.91 \pm 0.044 \pm 0.020 \pm 0.013, \\
& \rho^{2}=0.77 \pm 0.039 \pm 0.019 \pm 0.032
\end{align*}
$$

where the first error is statistical, the second MC statistical, and the third systematic.
These are to be compared with the previously published CLEO results [4] of $(1.18,0.72,0.92)$ with statistical errors of $(0.30,0.22,0.16)$ and systematic errors of $(0.12$, $0.07,0.06)$. It is clear that we have reduced the error on these extracted quantities significantly.

See Section 6.8.2 for goodness-of-fit test results on our extracted parameters vs. the CLEO ones.

## Bibliography

[1] Review of Particle Physics,, K.Hagiwara et al., Phys. Rev. D66, 010001 (2002), [http:/ /pdg.lbl.gov]
[2] Irish Scientists,"Lord Kelvin", [http://members.tripod.com/Ĩrishscientists/listof.htm]
[3] M.S. Gill, PhD Thesis, UCB, 2004.
[4] A. Ryd, PhD Thesis, UCSB, 1997; CLEO Collab., Phys. Rev. Lett. 76, 3898 (1996).
[5] C.M. LeClerc Ph.D.Thesis, UCB, Dec. 2002
[6] R.N. Cahn, 1996, Rev. Mod. Phys. Vol 68, pg 951.
[7] Peskin and Schroeder, "An Introduction to Quantum Field Theory", AddisonWestley Publishing, 1997.
[8] F. Halzen and A.D. Martin, "Quarks and Leptons", John Wiley and Sons, 1984.
[9] Y. Nir, "CP Violation In and Beyond the Standard Model", SLAC Summer Institute, 1999.
[10] P. Harrison and H.R. Quinn, "The BaBar Physics Book", SLAC, October 1998.
[11] [http://arxiv.org/abs/hep-lat/0309090], "Pentaquark hadrons from lattice QCD", F.Csikor, Z. Fodor, S.D. Katz, T.G. Kovacs, (2003).
[12] N. Isgur and M.B. Wise, Phys. Lett. B237, 527 (1990).
[13] N. Cabibbo, Phys. Rev. Lett. 10, 531 (1963); M. Kobayashi and T. Maskawa, Prog. Theor. Phys. 49, 652 (1973).
[14] H. Murayama, Massive Neutrino observation points at potential for RH neutrino existence
[15] "WEAK DECAYS OF HEAVY MESONS IN THE STATIC QUARK APPROXIMATION" N. Isgur, M. B. Wise, Phys.Lett.B232:113,1989
[16] M. Neubert, Physics Reports, 1994
[17] " QUARK MODEL FORM-FACTORS FOR HEAVY QUARK EFFECTIVE THEORY", F.Close, Wambach, Nucl.Phys.B412:169-180, (1994), (hep-ph/9307260)
[18] " HEAVY QUARK SYMMETRY IN B -i D(*) LEPTON ANTI-NEUTRINO SPECTRA", Z.Ligeti, B.Grinstein, Phys.Lett.B526:345-354 (2002), (hep-ph/0111392)
[19] " DISPERSIVE BOUNDS ON THE SHAPE OF ANTI-B -i D(*) LEPTON ANTI-NEUTRINO FORM-FACTORS", I. Caprini, L. Lellouch, M. Neubert, Nucl.Phys.B530:153-181,1998, (hep-ph/9712417)
[20] I. Caprini, M. Neubert, Phys. Lett. B 380, (1996) 376
[21] E687 Expt.; FNAL; [http://www.hep.uiuc.edu/e687]
[22] W.Verkerke, D.Kirkby, [http://roofit.sourceforge.net].
[23] R.Brun et.al. [http://root.cern.ch]
[24] BABAR Analysis Document 18: D. Kirkby et al, "A User's Guide to the RooFitTools Package for Unbinned Maximum Likelihood Fitting"
[25] F.James, [http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html]
[26] BABAR Analysis Document 34: C. Cheng, D. Kirkby, C. Roat, "Selection of D*lepton Events for B Lifetime and Mixing Analyses"
[27] BABAR Analysis Document 35: Bernard, Denis ; Bonneaud, Gerard R. ; Brochard, Francois ; Cohen-Tanugi, Johann ; Ferrag, Samir ; T’Jampens, Stephane ; Thiebaux, Christophe ; Vasileiadis, Georges ; Verderi, Marc; "Amplitude measurement in the B -] JPsi K* channels"
[28] BABAR Analysis Document 54: O. Long, "Slow pion efficiency"
[29] BABAR Analysis Document 90, Electron PID group, "Cut Based Electron Identification".
[30] BABAR Analysis Document 116, G. Mancinelli et al, "Kaon Selection at the BaBar experiment".
[31] BABAR Analysis Document 137: P. Burchat, C. Cheng, D. Kirkby, C. LeClerc, T.

Meyer, N. Roe, "Measurement of the $B^{0}$ Lifetime and Oscillation Frequency Using Fully-Reconstructed Semileptonic Decays"
[32] BABAR Analysis Document 147: P. Burchat, C. Cheng, D. Kirkby, T. Meyer, "Semileptonic B Sample for Lifetime and Mixing Analyses"
[33] BABAR Analysis Document 154: Bernard, Denis ; Bonneaud, Gerard R. ; Brochard, Francois ; Cohen-Tanugi, Johann ; Ferrag, Samir ; T'Jampens, Stephane ; Thiebaux, Christophe ; Vasileiadis, Georges ; Verderi, Marc " Measurement of the B -] J/Psi K* decay amplitudes"
[34] BABAR Analysis Document 324: Bona, Marcella ; Cavoto, Gianluca ; Dallapiccola, Carlo ; Grancagnolo, Sergio ; Varnes, Erich ; Wagner, Stephen Robert ; Zhang, Lei "Report of the Tracking Efficiency Task Force"
[35] PID AWG Recommendations; [http:/ /www.slac.stanford.edu/BFROOT/www/Physics/Tools/Pid,
[36] BABAR Analysis Document 457: Borean, Cristiano ; De Nardo, Guglielmo ; Della Ricca, Giuseppe ; Monorchio, Diego ; Rotondo, Marcello, "Measurement of the $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ Branching Ratio using the 2001 Data Sample"
[37] BABAR Analysis Document 496: Snyder, Art; "CP violation in $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ Decay"
[38] BABAR Analysis Document 531: Weinstein, Amanda; "Measurement of Exclusive Charmless Semileptonic Decays", Dec. 2003
[39] BABAR Analysis Document 532: Rotondo, Marcello et.al.; "Determination of $\left|V_{c b}\right|$ from the decay $\bar{B}^{0} \rightarrow D^{*+} \ell^{-} \bar{\nu}^{\prime \prime}$, May 2004
[40] BABAR Analysis Document 414 : Mazur, Michael; " $B^{0} \rightarrow D^{*} \tau \nu$ Decay"
[41] BABAR Analysis Document 792: Paoloni, Eugenio et.al. "A search of T-violation in semileptonic B decays: feasibility study."
[42] BABAR Analysis Document 527: Dubitzky, Rolf; "Measurement of the CKM Matrix Element $V_{c b}$ with $B^{0} \rightarrow D^{*-} e^{+} \nu$ Decays"
[43] BABAR Analysis Document 664: Fortin, Dominique et. al. ; "Measurement of —Vub—from semileptonic B decays using the lepton energy and q2"
[44] H. Chernoff, E. Lehmann, Annals of Stats., 1954
[45] Grancagnolo, Sergio; Helicity method of evaluating slow pion efficiency; [http://www.slac.stanford.edu/grancagn/internal/softpion/datamc/nccweightedmc.html]
[46] "Measurement of $B \rightarrow D$ Form Factors in the Semileptonic Decay $B^{0} \rightarrow D^{*} \ell \nu$ at BaBar", ICHEP 2004, BABAR Collaboration

## Appendix A

## Notation Used in Text

## A. 1 Colors

Though a color version (either printed or electronic) of this document is not necessary to interpret the figures and formulae within, we have found that color adds to the ease of visually inspecting components of some of them, and for reference we collect here the notation of colors used (terminology described in body as colors are introduced):

- Purple is used for kinematic variables .
- Red is used for form factors .
- Green is used for form factor ratios and the form factor slope $\rho^{2}$.


## A. 2 Standard Form Factor Parameter Notation

For quick reference, the legend followed in the text for EvtGen shorthand parameter notation (this is condensed from Sec.7.2.1) is the following: "XYZ" refers to the value of $\left(R_{1} ; R_{2} ; \rho^{2}\right)$ respectively. The letter codes are these: $\mathrm{A}=0.50, \mathrm{~B}=0.75, \mathrm{C}=1.00$, $\mathrm{D}=1.25, \mathrm{E}=1.50$, and L are the CLEO values of $(1.18 ; 0.72 ; 0.92)$. So e.g. the set labelled "LLL" would mean $R_{1}=1.18, R_{2}=0.72, \rho^{2}=0.92$, while " ABE " would mean $R_{1}=0.50$, $R_{2}=0.75, \rho^{2}=1.50$, etc.

We denote this "Standard Form Factor Parameter Notation".

## Appendix B

## Including CP violation and massive

## lepton terms in Fitting PDF

Here we discuss what would happen in the case were CP violation not to be disallowed in this decay ab initio, and also how much impact the terms of the PDF which contain a massive lepton can have. If we allow for strong phases in the decay then we cannot assume that the helicity amplitudes are real (see [37]). Further, for non-negligible lepton mass there is an extra transverse helicity amplitude $H_{t}$. Generalizing to allow these two possibilities ( CP violation and nonzero lepton mass), we obtain for the full PDF the lengthy expression (from [37]):

$$
\frac{d \Gamma}{d q^{2} d \cos \theta_{\ell} d \chi d \cos \theta_{V}}=\frac{3 G_{F}^{2}\left|V_{c b}\right|^{2}}{8(4 \pi)^{4}} \frac{\wp_{D^{*}}}{\left(q^{2}-m_{\ell}^{2}\right)^{2}} q^{2} B\left(D^{+*} \rightarrow D^{0} \pi^{+}\right)
$$

$$
\begin{gathered}
\times\left[\left(\left(1+\cos ^{2} \theta_{\ell}\right)+\frac{m_{\ell}^{2}}{q^{2}} \sin ^{2} \theta_{\ell}\right) \sin ^{2} \theta_{V}\left(\left|H_{+}\right|^{2}+\left|H_{-}\right|^{2}\right)\right. \\
+4\left(\sin ^{2} \theta_{\ell}+\frac{m_{\ell}^{2}}{q^{2}} \cos ^{2} \theta_{\ell}\right) \cos ^{2} \theta_{V}\left|H_{0}\right|^{2} \\
-2 \cos \theta_{\ell} \sin ^{2} \theta_{V}\left(\left|H_{+}\right|^{2}-\left|H_{-}\right|^{2}\right) \\
-2\left(1-\frac{m_{\ell}^{2}}{q^{2}}\right) \sin ^{2} \theta_{\ell} \cos 2 \chi \sin ^{2} \theta_{V} R e\left(H_{+} H_{-}^{*}\right) \\
+\left(1-\frac{m_{\ell}^{2}}{q^{2}}\right) \sin 2 \theta_{\ell} \cos \chi \sin 2 \theta_{V} R e\left(\left(H_{+}+H_{-}\right) H_{0}^{*}\right) \\
\left.-2 \sin \theta_{\ell} \cos \chi \sin 2 \theta_{V} R e\left(H_{+}-H_{-}\right) H_{0}^{*}\right) \\
+4 \frac{m_{\ell}^{2}}{q^{2}} \cos ^{2} \theta_{V}\left|H_{t}\right|^{2} \\
-8 \frac{m_{\ell}^{2}}{q^{2}} \cos \theta_{\ell} \cos ^{2} \theta_{V} R e\left(H_{t} H_{0}^{*}\right) \\
+2 \frac{m_{\ell}^{2}}{q^{2}} \sin \theta_{\ell} \cos \chi \sin 2 \theta_{V} R e\left(\left(H_{+}+H_{-}\right) H_{t}^{*}\right) \\
+2\left(1-\frac{m_{\ell}^{2}}{q^{2}}\right) \sin \theta_{\ell} \theta_{\ell} \sin 2 \chi \sin \theta_{V} \operatorname{Im}\left(H_{+} H_{-}^{*}\right) \\
-\left(1-\frac{m_{\ell}^{2}}{q^{2}}\right) \sin 2 \theta_{\ell} \sin \chi \sin 2 \theta_{V} \operatorname{Im}\left(\left(H_{+}-H_{-}\right) H_{0}^{*}\right) \\
+2 \sin \theta_{\ell} \sin \chi \sin 2 \theta_{V} \operatorname{Im}\left(\left(H_{+}+H_{-}\right) H_{0}^{*}\right)
\end{gathered}
$$

This is the form which shows us both the terms that would be different if this decay did not conserve CP , and if the mass of the lepton played a significant role. As for the former, there is thus far no evidence that CP is not conserved in this decay, both from theoretical considerations (any CP violation would have to derive from fairly exotic beyond the SM physics, see e.g. [37] and references therein), or from direct observation at BaBar thus far [41].

For the lepton mass: note all appearances of the transverse helicity amplitude $H_{t}$ come in with a factor of $m_{\ell}$ which indicates that this term can be neglected if $m_{\ell}^{2} / q^{2}$ is a negligible quantity.

However the terms with $m_{\ell}$ factors do need to be kept for $B^{0} \rightarrow D^{*} \tau \nu$ decay such as that being measured in [40].

Further, to find out how much the actual light lepton masses impact the analysis, we may do a simple calculation first of all which shows that if the events were to obey a flat distribution over the entire allowed $q^{2}$ region, then the place where a nonzero $m_{\ell}^{2}$ would be significant would be in the region where $m_{\ell}^{2} / q^{2}>0.1$, which happens for the muon mass only in the region of $q^{2}$ almost at its minimal value, which is approximately $0.001 \%$ of the entire allowed $q^{2}$ region, indicating the mass is not likely a large effect.

To do this more accurately we can take phase space on $q^{2}$ into account to see how many events would actually fall into this region, and just by eye we can see that the $q^{2}$ plot falls off at both edges, in particular here at the minimal $q^{2}$ (maximal $w$ ) region, so that this can only further suppress the effect.

Since the electron mass is over 200 times smaller than the muon mass, the above conclusion will apply even more fully to the electron in the event.

## Appendix C

## EvtGen and RooFit PDF Validation

## C. 1 EvtGen and RooFit Validation from Kinematic Variable Distribution Comparison

In this Appendix we will be showing some results from EvtGen vs. RooFit validation tests, and also describing how the ultimate root files we did our analysis on were generated.

See Section 5.1.1 for a description of the EvtGen MC generation package.
RooFit[22] is the fitting package built on Root[23] which we use to make plots and encode the PDF to which we fit the data. It uses as its fitting engine the standard Minuit program[25].

Figure C. 1 is a plot of the the projections of the four kinematic variables from two different sources: the curve is from the signal PDF encoded into RooFit and then
having a number of events thrown with this PDF and binned, and the histograms with statistical error bars are from 100k events thrown in the same PDF as encoded into EvtGen. We include into EvtGen the same parameters as are in the RooFit PDF. Neither PDF has any detector acceptance or any other effects on any of the tracks.

In the EvtGen sample we force the $B^{0}$ to decay purely to a $D^{*} \ell \nu$ event (this $B R$ in actuality is about $11 \%$ in total counting both light lepton modes), and we take the lepton to be only an $e^{-}$. The only place at the EvtGen level that the flavor of the lepton can matter is with respect to final state radiation (FSR) in which a photon is radiated off of an lepton leg internally in the Feynman diagram of the process (as opposed to Bremsstrahlung which is a separate external process after all the particles have exited the Feynman diagram, and which takes place when the leptons interact with the protons of the detector material as they pass through the layers of the detector). FSR effects are probed in detail in Section 8.4.5, where it can be seen that their ultimate affect on this analysis is quite small).

The $D^{*}$ is further forced to decay to a $D^{0}$ and a (slow) $\pi$ (in real life this has a $\frac{2}{3}$ branching ratio). The $\pi$ from the $D^{*}$ is normally termed "slow" (and often denoted $\pi_{s}$ ) because there is so little phase space allowed for this decay (i.e., the mass of the $D^{*}$ nearly equals the sum of the masses of the $D^{0}$ and a charged pion), that the momentum of the resultant $\pi$ is in general very low (in fact, in the $B^{0}$ rest frame it varies from 0-250 $\mathrm{MeV} / \mathrm{c}$ or so). We then calculate the event level kinematic variable quantities from the event level (interaction point) particle four-vectors of the $D^{0}, \pi_{s}, \ell$ (in a real reconstruction
the measured tracks in the detector are a $K, \pi, \pi_{s}, \ell$, and the energy of the $\pi^{0}$ daughter photons, for the $D^{0} \rightarrow K \pi \pi^{0}$ decay).

Note we also force in EvtGen the $D^{0}$ to decay only into the $K \pi$ mode (this has no effect on the kinematic variable distributions, but only second order effects on e.g. the acceptance through the multiplicity which affects the track-finding of the event etc., which is discussed in Section 8.4.1).

The overlap in the curves and the histograms in Figure C. 1 to C. 6 simply shows that the PDF's in the two places match, and thus our EvtGen PDF is exactly that which we will be fitting with with our fitting tools (the labelling of the parameter sets in the captions follows the convention listed at the end of Section 7.2.1). It also shows that we know exactly how to control the three fit parameters which are used as inputs in EvtGen production, as the plots are not identical for different input parameters. Though a trivial consistency check, this is foundational for all future work, and is useful to do before proceeding onward for such a multivariate and lengthy PDF.

## C. 2 Showing Equivalence of PDF by Fitting

Another way to see that we are using identically equivalent forms for the theoretical PDF in the fitting code and in the generation code is to fit an EvtGen sample with the theoretical PDF normalized by simply dividing through by the factor we obtain by integrating the PDF expression over the entire allowed 4-dimensional kinematic variable space. For completeness, we give this normalization here (where we have taken eq.(3.5),


Figure C.1: Overlaying of generated MC (EvtGen, points) and fitting PDF (RooFit, curve), both with LLL parameters.
substituted in eq.(3.8) and integrated over the 4-dimensional kinematic variable space) :
$\iiint \int_{4 \mathrm{kv} v^{\prime}} g^{[t h]}=$
$11.90848681-7.734921707 \rho^{2}+1.449766988 \rho^{4}$
$+.2849279646 R_{1}^{2}-.1637048251 R_{1}^{2} \rho^{2}+.02670941074 R_{1}^{2} \rho^{4}$
$-6.422176045 R_{2}+4.908824973 \rho^{2} R_{2}-.998749412 \rho^{4} R_{2}$
$+1.204216628 R_{2}^{2}-.962930182 \rho^{2} R_{2}^{2}+.201215453 \rho^{4} R_{2}^{2}$


Figure C.2: Overlaying generated MC (EvtGen, red crosses) and fitter MC (RooFit Toy MC, black histograms with error bars ) both with LLL parameters (note that the RooFit "curve" in this case and for all Figures C.2-C. 6 have been generated by throwing events with the theoretical PDF and binning them, not by taking a projection of the PDF itself as in Figure C.1, which is why the "curves" appear jagged).

We now take samples generated from eq.(3.5) with various combinations of the parameters, and fit to it with the same PDF (normalized with the above expression) and find the results in Tables C. 1 to C.6).

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.183 | 0.011 |
| $R_{2}$ | 0.720 | 0.713 | 0.013 |
| $\rho^{2}$ | 0.920 | 0.923 | 0.011 |

Table C.1: LLL file, 100k events - see 7.2.1 for parameter labelling nomenclature

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 0.500 | 0.512 | 0.010 |
| $R_{2}$ | 0.500 | 0.503 | 0.014 |
| $\rho^{2}$ | 0.500 | 0.494 | 0.015 |

Table C.2: AAA file, 100k events

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.500 | 1.488 | 0.010 |
| $R_{2}$ | 1.500 | 1.498 | 0.009 |
| $\rho^{2}$ | 1.500 | 1.494 | 0.007 |

Table C.3: EEE file, 100k events

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.000 | 0.972 | 0.031 |
| $R_{2}$ | 1.000 | 1.001 | 0.034 |
| $\rho^{2}$ | 1.000 | 0.976 | 0.033 |

Table C.4: CCC file, 10k events


Figure C.3: Overlaying of generated MC (EvtGen, crosses) and fitter MC (RooFit Toy MC, histograms with error bars ) both with LLL parameters; EvtGen has FSR here (only in this case).

## C. 3 Ntuple production

The DstarlnuUser package is used to produce HBOOK files for this analysis. The events are filtered from the "very tight" tag-bits that were written to the tag data base during skimming. For these events, the candidate lists are regenerated. The output of each job is an HBOOK file containing an ntuple.

Each ntuple is processed into a form readable by ROOT (with h2root), further


Figure C.4: Overlaying of generated MC (EvtGen, crosses) and fitter MC (RooFit Toy MC, histograms with error bars ), both with CCC parameters.
processed by C++ code into final small consolidated ROOT files, then a final selection is run in RooFit that applies the ultimate selection cuts. In the case of multiple $B^{0} \rightarrow D^{*} \ell \nu$ candidates with the same $D^{0}$ decay (but different $D^{0}$ candidates), the $D^{0}$ candidate with reconstructed mass, based on simple four-vector addition, closest to the nominal mass is selected. If multiple $B^{0} \rightarrow D^{*} \ell \nu$ candidates overlap but use different $D^{0}$ decay modes, the candidates are retained. If there are multiple $B^{0} \rightarrow D^{*} \ell \nu$ candidates with the same $D^{0}$ candidate, only the first candidate in the list is kept.


Figure C.5: Overlaying of LLL generated MC (EvtGen, crosses) and CCC fitter MC (RooFit Toy MC, histograms with error bars ).

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.000 | 1.004 | 0.010 |
| $R_{2}$ | 1.000 | 0.994 | 0.011 |
| $\rho^{2}$ | 1.000 | 0.996 | 0.010 |

Table C.5: CCC file, 100k events


Figure C.6: Overlaying of generated (EvtGen, histogram) and fitting (RooFit, curve) PDF's - both with LLL parameters.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.000 | 0.9968 | 0.0031 |
| $R_{2}$ | 1.000 | 1.0032 | 0.0034 |
| $\rho^{2}$ | 1.000 | 0.9989 | 0.0032 |

Table C.6: CCC file, 1M events

## Appendix D

## EvtGen Moments Method Validation

In this Appendix we show generator-level MC tests of the Moments Fitting Method.

## D. 1 Fix calibration on one file and fit to files with varying parameters

As an initial test of the Moments Method, we calibrate on a pure signal EvtGen file with all R parameters fixed at 1.00, and fit to statistically independent files generated with R parameters also all fixed at 1.00 (to see details on how events were generated in this EvtGen sample, see Section 5.1.1).

Note that by "calibration", we here mean finding the acceptance correction function via the $18 \xi$ values (the integrated efficiency moments) from a given sample, as discussed in the previous sections.

Each file is about 100k events, pure signal mode (still $e^{-}$only, $D^{0} \rightarrow K \pi$ mode), with no background and no smearing. All kinematic variables are evaluated in the $B^{0}$ rest frame, since we can go to that frame exactly in MC.

We first take a sample with no cuts, see Table D. 1 for fit results (inputs are 1.00 for each of the three parameters).

Note in the first line the level to which the fit gives back the input MC parameters - this is precisely a result of the theorem of Section 6.2.5, and thus demonstrates the consistency of the method quite clearly at this step.

| Sample Number | $R_{1}$ Fitval | $R_{1}$ Error | $R_{2}$ Fitval | $R_{2}$ Error | $\rho^{2}$ Fitval | $\rho^{2}$ Error |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1 calib. sample | 1.0000 | 0.01 | 1.0000 | .01 | 1.0000 | .01 |
| 2 | 1.00 | 0.01 | 1.01 | .01 | 0.99 | .01 |
| 3 | 1.00 | 0.01 | 1.01 | .01 | 0.99 | .01 |
| 4 | 1.00 | 0.01 | 1.01 | .01 | 0.99 | .01 |
| 5 | 1.00 | 0.01 | 1.02 | .01 | 0.98 | .01 |

Table D.1: Calibrating on one EvtGen sample, then fitting to five statistically independent EvtGen samples with no cuts. Inputs are $R_{1}=R_{2}=\rho^{2}=1.00$.

## D. 2 First test of acceptance correction method with cuts

We now apply the following cuts to both calibration and fit samples: $p_{l}>$ $1.2 \mathrm{GeV}, p_{\pi_{s}}>65 \mathrm{MeV}, 155^{\circ}>\theta_{l}>25^{\circ}$ (lab lepton momentum, lab slow pion momentum, and lab lepton polar angle with respect to the detector axis). These are the cuts we have found that most strongly affect the relevant kinematic variable distributions, so it is reasonable to apply only them to the EvtGen files, though for this test, any cuts will do

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.858 | .012 |
| $R_{2}$ | 0.720 | 0.987 | .011 |
| $\rho^{2}$ | 0.920 | 0.938 | .010 |

Table D.2: Analytic calibration, fit to 111 k file with cuts.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.156 | .014 |
| $R_{2}$ | 0.720 | 0.719 | .014 |
| $\rho^{2}$ | 0.920 | 0.920 | .011 |

Table D.3: Moment method calibration from a large (1M) file with cuts in it, fit to same 111k file with cuts.
as long as they are applied equivalently to the calibration and fit samples.
What we first do here is take a 200k event sample that is reduced to about 112 k events after the above cuts, and fit it with the analytic normalization as done in Section C.2. The results are shown in Table D.2, and it can be seen how dramatically they diverge from the inputs.

We then apply the acceptance correction and re-fit with this calibration to obtain the results of Table D.3. Now it is seen that the method has brought the results within $1-2 \sigma$ of the original inputs, thus, this is a first demonstration of the method working correctly.

## D. 3 Tests varying fit samples

Following a similar fitting procedure to Section D.1, but now iterating over fit and calibration files with cuts in them, yields the fit results in Table D.4. Note again in the first line the level to which the fit gives back the input MC parameters as it must according to the self-calibration theorem.

| Sample Number | $R_{1}$ Fitval | $R_{1}$ Error | $R_{2}$ Fitval | $R_{2}$ Error | $\rho^{2}$ Fitval | $\rho^{2}$ Error |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1 calib. sample | 1.0000 | 0.02 | 1.0000 | .02 | 1.0000 | .01 |
| 2 | 1.00 | 0.02 | 1.00 | .02 | 0.99 | .01 |
| 3 | 1.00 | 0.02 | 1.01 | .01 | 0.99 | .01 |
| 4 | 1.00 | 0.02 | 1.02 | .01 | 0.98 | .014 |
| 5 | 0.99 | 0.02 | 1.03 | .015 | 0.96 | .014 |

Table D.4: Calibrating on one EvtGen sample, then fitting to five statistically independent EvtGen samples with cuts as specified in the text. Inputs are $R_{1}=R_{2}=\rho^{2}=1.00$.

In order to confirm that the fit tests above are not unduly sensitive to the input values specified for the fit file, we do another test where we calibration on a file with parameters fixed at 1.00, but the input parameters vary in the fit files as listed (to be more realistic, these files have cuts, again). In this case, we choose to calibration on a large sample with 1 M events, so as not to introduce any spurious error due to the calibration file size (see Section 8.2.2 for a discussion of this error). See Table D. 5 for results.

Finally, we do a test to make sure we are not sensitive to the values chosen to calibration the sample, Table D. 6 shows results of this test.

Since the results of all these tests show that nearly all fits give back the input pa-

| Sample Number | $R_{1}$ in | $R_{1}$ fit | $R_{2}$ in | $R_{2}$ fit | $\rho^{2}$ in | $\rho^{2}$ fit |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1 calib. sample | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | 0.50 | $0.51 \pm .01$ | 0.50 | $0.51 \pm .02$ | 0.50 | $0.50 \pm .02$ |
| 3 | 0.75 | $0.75 \pm .01$ | 0.75 | $0.76 \pm .02$ | 0.75 | $0.76 \pm .02$ |
| 4 | 1.25 | $1.27 \pm .015$ | 1.25 | $1.24 \pm .01$ | 1.25 | $1.26 \pm .01$ |
| 5 | 1.50 | $1.49 \pm .01$ | 1.50 | $1.52 \pm .013$ | 1.50 | $1.48 \pm .01$ |

Table D.5: Calibrating on one EvtGen sample of 1 M events generated from input values $R_{1}=R_{2}=\rho^{2}=1.00$, then fitting to four statistically independent EvtGen samples with varying parameters, and cuts as specified in the text. Minuit statistical errors shown with the fit results.

| Calib Sample Number | Calib File Parameters | $R_{1}$ Fitval | $R_{2}$ Fitval | $\rho^{2}$ Fitval |
| :--- | :--- | :---: | :---: | :---: |
| 1 | $(0.50 ; 0.50 ; 0.50)$ | $0.99 \pm .01$ | $1.02 \pm .015$ | $0.98 \pm .015$ |
| 2 | $(0.75 ; 0.75 ; 0.75)$ | $1.00 \pm .01$ | $1.02 \pm .015$ | $0.98 \pm .015$ |
| 3 | $(1.25 ; 1.25 ; 1.25)$ | $1.00 \pm .01$ | $1.02 \pm .015$ | $0.99 \pm .015$ |
| 4 | $(1.50 ; 1.50 ; 1.50)$ | $0.99 \pm .01$ | $1.02 \pm .015$ | $0.99 \pm .015$ |

Table D.6: Calibrating on multiple EvtGen samples of 1M events generated with input parameter values as shown, then fitting to an EvtGen sample with input values $R_{1}=$ $R_{2}=\rho^{2}=1.00$. All files have cuts as specified in the text. Minuit errors shown with the fit results.
rameters within one or two $\sigma$, we may conclude that we can trust our fitting procedure on at least EvtGen generated events, and we assign no systematic error in the intrinsic method from what parameter set was used to do the calibration vs. the ultimate fit parameters.

## Appendix E

## SP4 Fits by separate $D^{0}$ and lepton

## modes

While the fit tests in 6.3.1 on SP4 MC were been shown in aggregate for both the calibration and fit files - that is, with the two lepton modes and four $D^{0}$ decay modes grouped all together - it is useful to check how fits fare when the calibration and fit files are specific to each of these modes. That is, one might expect that the acceptance correction would work better when we specifically calibration on the file with the lepton and $D^{0}$ mode isolated, and fit to a sample with the same isolated lepton and $D^{0}$ mode, because of small differences in tracking, PID etc. (i.e., collectively: acceptance) on the daughter tracks between these different modes.

Since we work with four $D^{0}$ decay mode modes and two lepton modes, this gives us eight individual samples we could perform our tests on (though since the $D^{0} \rightarrow$
$K_{S} \pi \pi$ mode has such low statistics relative to the primary three, we chose to do these tests leaving this mode out as it will not have a significant impact on the total fit in any case).

With our initial relatively small size sample MC we did this test for now just separating out the two lepton modes (with the $D^{0}$ modes lumped together), then the main three $D^{0}$ modes (with the leptons lumped together), because this should still give us some improvement with respect to lumping them all together in any case.

Initial results indicate this is indeed mildly the case for the $D^{0}$ modes (this can be seen by eye in Table E.2-E. 4 - note we chose to use only 3 significant figures in the fit results given here, since the statistical errors are so large on these samples), where the fits on a given $D^{0}$ mode are closer when calibrated on a sample composed purely of that $D^{0}$ mode.

It does not appear that separating by lepton mode gives any obvious improvement in the fit (see Table E.1) - still, because our calibration method works best for calibration samples that are most like the data we eventually fit, it intuitively makes the most sense to separate all modes out by both lepton and $D^{0}$ decay mode, which is what we do in the end.

| Param | Input | Fit Value | Error | Param | Input | Fit Value | Error |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.188 | .058 | $R_{1}$ | 1.180 | 1.267 | .061 |
| $R_{2}$ | 0.720 | 0.632 | .048 | $R_{2}$ | 0.720 | 0.712 | .046 |
| $\rho^{2}$ | 0.920 | 0.947 | .037 | $\rho^{2}$ | 0.920 | 0.937 | .037 |
| Param | Input | Fit Value | Error | $\overline{\text { Param }}$ | Para | Input | Fit Value |
| Error |  |  |  |  |  |  |  |
| $R_{1}$ | 1.180 | 1.218 | .064 | $\mid R_{1}$ | 1.180 | 1.332 | .067 |
| $R_{2}$ | 0.720 | 0.657 | .053 | $R_{2}$ | 0.720 | 0.784 | .049 |
| $\rho^{2}$ | 0.920 | 0.951 | .039 | $\rho^{2}$ | 0.920 | 0.911 | .040 |
| Param | Input | Fit Value | Error | $\overline{\mid \text { Param }}$ | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.104 | .059 | $R_{1}$ | 1.180 | 1.197 | .062 |
| $R_{2}$ | 0.720 | 0.669 | .044 | $R_{2}$ | 0.720 | 0.711 | .044 |
| $\rho^{2}$ | 0.920 | 0.912 | .037 | $\rho^{2}$ | 0.920 | 0.925 | .037 |

Table E.1: Calibration on various lepton files: Fits on left are to electron-only file, on right to muon only file. Calibration from top to bottom: on file with both leptons, on electrononly file, and on a muon only file (note the fit and calibration files are only on partially different statistically independent files in this case). The errors are the Minuit-quoted statistical errors only, for this and all following tables in this Appendix.

| Param | Input | Fit Value | Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.12 | .05 |
| $R_{2}$ | 0.720 | 0.76 | .04 |
| $\rho^{2}$ | 0.920 | 0.98 | .03 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.15 | .05 |
| $R_{2}$ | 0.720 | 0.75 | .05 |
| $\rho^{2}$ | 0.920 | 0.88 | .04 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.08 | .05 |
| $R_{2}$ | 0.720 | 0.79 | .04 |
| $\rho^{2}$ | 0.920 | 1.07 | .03 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.14 | .05 |
| $R_{2}$ | 0.720 | 0.74 | .04 |
| $\rho^{2}$ | 0.920 | 1.00 | .03 |

Table E.2: All fits to mode $1(K \pi)$, calibration on various $D^{0}$ decay mode files. Calibration from top to bottom: all modes combined, mode 1, 2, and 3 ( $1=K \pi, 2=K 3 \pi, 3=K \pi \pi^{0}$ ) (note the fit and calibration files are only on partially different statistically independent files in this case).

| Param | Input | Fit Value | Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.19 | .06 |
| $R_{2}$ | 0.720 | 0.72 | .05 |
| $\rho^{2}$ | 0.920 | 0.84 | .04 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.22 | .06 |
| $R_{2}$ | 0.720 | 0.71 | .05 |
| $\rho^{2}$ | 0.920 | 0.72 | .04 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.16 | .06 |
| $R_{2}$ | 0.720 | 0.74 | .04 |
| $\rho^{2}$ | 0.920 | 0.95 | .04 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.21 | .06 |
| $R_{2}$ | 0.720 | 0.70 | .04 |
| $\rho^{2}$ | 0.920 | 0.85 | .04 |

Table E.3: All fits to mode $2(K 3 \pi)$,calibration on various $D^{0}$ decay mode files. Calibration from top to bottom: all modes combined, mode 1,2 , and $3(1=K \pi, 2=K 3 \pi$, $3=K \pi \pi^{0}$ ). (note the fit and calibration files are only on partially different statistically independent files in this case).

| Param | Input | Fit Value | Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.180 | 1.13 | .06 |
| $R_{2}$ | 0.720 | 0.76 | .05 |
| $\rho^{2}$ | 0.920 | 0.90 | .04 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.15 | .06 |
| $R_{2}$ | 0.720 | 0.76 | .05 |
| $\rho^{2}$ | 0.920 | 0.79 | .05 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.09 | .06 |
| $R_{2}$ | 0.720 | 0.78 | .05 |
| $\rho^{2}$ | 0.920 | 1.00 | .04 |
| Param | Input | Fit Value | Error |
| $R_{1}$ | 1.180 | 1.15 | .06 |
| $R_{2}$ | 0.720 | 0.73 | .04 |
| $\rho^{2}$ | 0.920 | 0.92 | .04 |

Table E.4: All fits to mode $3\left(K \pi \pi^{0}\right)$, calibration on various $D^{0}$ decay mode files . Calibration from top to bottom: all modes combined, mode 1,2 , and $3(1=K \pi, 2=K 3 \pi$, $3=K \pi \pi^{0}$ ). (note the fit and calibration files are only on partially different statistically independent files in this case).

## Appendix F

## YFrame

In past analyses [4], the "Y-average frame" (or Y-frame for short) was that frame in which the resolution of the kinematic variables was smallest, and thus it gave the smallest uncertainty relative to what the kinematic variables in the true $B^{0}$ rest frame would be. Since the $B^{0}$ momentum relative to the $\Upsilon(4 \mathrm{~S})$ rest frame is nominally just 327 $\mathrm{MeV} / \mathrm{c}$ (and the stable particles from this decay are in the few- $\mathrm{GeV} / c$ range), the change in the kinematic variables is not large, nonetheless, we will see it is large enough that the resolution is significantly compromised (in fact, this is the largest source of resolution for the kinematic variables, compared to the detector-smearing contributions etc.).

In the crudest $\Upsilon(4 \mathrm{~S})$ approach, we take the $\Upsilon(4 \mathrm{~S})$ rest frame to be the $B^{0}$ rest frame, and then calculate the kinematic variables, and subtract the MC truth values (this is done on a pure EvtGen level MC, so no smearing or other effects intrude).

For the Y-frame on the other hand, we attempt a better approximation of the $B^{0}$
frame using some information known to us: the average momentum magnitude of the $B^{0}$ in the $\Upsilon(4 \mathrm{~S})$ frame, and the included angle between the $B^{0}$ and the pseudo-particle made up by summing the 4 -momenta of the $D^{*}$ and lepton in the $\Upsilon(4 \mathrm{~S})$ frame. The included angle between the $B$ and the pseudoparticle " $Y$ "is given by eq. (5.7).

For the Y -frame, we first take the $B^{0}$ to lie on a cone with the known Y system vector as its axis, and opening angle $\cos \theta_{B Y}$. We then take the plane formed by the Y and $D^{*}$ vectors, which intersects the cone along two lines, and assume that the $B^{0}$ lies along first one, then the other of these two lines, calculating the four kinematic variables in each case, and taking the average of the two values for the final result.

To isolate and observe the Y-frame effect, we take the Y-frame-calculated kinematic variables, and again subtract the MC-truth values. For this test, this is done on a pure EvtGen level MC, so no smearing or other effects intrude.

In Figure F.1, we see the result of this procedure and the comparison of the resolutions obtained in the two approaches, and it is clearly seen that the resolution is significantly better in the Y-frame approach compared to the the aforementioned $\Upsilon(4 S)$ based approach.

Note that though for a fixed rest mass of the $\Upsilon(4 S)$ the momentum of each daughter $B^{0}$ is fixed at $327 \mathrm{MeV} / c$, there are two things that lead to variation in the actual magnitude of the $B^{0}$ momentum: the intrinsic BaBar beam energy beam spread (beamwidth) of some $5 \mathrm{MeV} / c$, and the intrinsic width of the $\Upsilon(4 S)$ of about $10 \mathrm{MeV} / c$. These lead to a variation in $p_{B^{0}}$ from about $287 \mathrm{MeV} / c$ to $369 \mathrm{MeV} / c$, and since the $\Upsilon(4 S)$


Figure F.1: $\Upsilon(4 S)$-based Kinematic Variable resolutions are in black, Y-frame resolutions in blue (dashed), top: left- $w$; right- $\cos \theta_{\ell} ;$ bottom: left- $\cos \theta_{V} ;$ right- $\chi$. In all cases the Y-frame resolutions are better, for $\cos \theta_{\ell}$ and $\cos \theta_{V}$ dramatically so (note that because the samples had the same number of events, all blue curves have been scaled down to fit within the frame).
width is twice as large, when convoluted with the beamwidth, the resultant $\mathrm{CM} p_{B^{0}}$ width is primarily given by the intrinsic beamwidth. The full MC properly models these widths and the $B^{0}$ momentum variation, and thus this variation in the magnitude does not then have to be separately added by hand.

## Appendix G

## Avoiding Spurious PDF Zeros for

## Background by PDF $\chi$-Averaging

Very few data points will be found by definition at places in the phase space at which the PDF goes to zero. However - there is no such limitation on the number of background events that may land at these points. Because of this, when the logarithm of the PDF is evaluated for these points as part of the log-likelihood computation, it will develop spuriously large (negative) values, driving the fit far astray from the actual minima.

The technique we developed to deal with these spurious zeros was to evaluate the PDF for all points (including the data, MC-selected background, and calibration samples, for consistency) at values of the phase space not exactly at the given point, but offset from it on both sides in $\chi$ by $\pm 0.1$ ( $\chi$ is chosen because it is at particular discrete values
of $\chi$ that the theoretical PDF can be seen to go to zero). We then took the values returned of the PDF for these two points and averaged them, taking this value as the actual value for the given point.

This in fact took care of all spurious zeros, and doing tests showed it did not bias the fit values, nor expand any of the errors. Varying the value by which we shifted $\chi$ showed that 0.1 was optimal in terms of being small enough to not bias the fit values, and large enough to eliminate spurious zero problems fully, so we chose to evaluate all fit results with this offset.

## Appendix H

## EvtGen Validation of DUBS Method

To test the DUBS method in the simplest case without resolution effects of real MC etc., we begin by doing the following: we calibrate on a sample of 100 k Evtgen events generated with the parameters all set to 1.00 , no cuts, and fit to a statistically independent sample of the same type. This fit yields the results in Table H.1.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.00 | 1.00 | .01 |
| $R_{2}$ | 1.00 | 0.99 | .01 |
| $\rho^{2}$ | 1.00 | 1.01 | .01 |

Table H.1: Calibration on Evtgen sample with all $\mathrm{R}=1.00$, fit to statistically independent sample of the same type.

We now add to this sample 20k events with the $R$ parameters set each to 0.50 , and re-fit, the results are in Table H.2.

The results are for each of the fitted parameters to come out about $20 \%$ biased
towards 0.50 as might be naively expected.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.00 | 0.93 | .01 |
| $R_{2}$ | 1.00 | 0.91 | .01 |
| $\rho^{2}$ | 1.00 | 0.94 | .01 |

Table H.2: Calibration on Evtgen sample with all $\mathrm{R}=1.00$, fit to statistically independent sample of the same type with 20 k of added events with all $\mathrm{R}=0.50$, fit is very off as expected.

Now subtract from this a statistically independent sample of the same type, i.e., 20k events with the R parameters set each to 0.50 , and redo the fit, the results are in Table H.3. It can be seen clearly that our first simplistic test of the DUBS method passes easily.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.00 | 1.01 | .01 |
| $R_{2}$ | 1.00 | 0.99 | .01 |
| $\rho^{2}$ | 1.00 | 1.01 | .01 |

Table H.3: Calibration on Evtgen sample with all $\mathrm{R}=1.00$, fit to statistically independent sample of the same type with 20k of added events with all $\mathrm{R}=0.50$, and 20 k statistically independent events with all $\mathrm{R}=0.50$ also subtracted off - fit is spot on.

## Appendix I

## Examples of need for bias map

## correction

In this Appendix we give two examples that demonstrate why we need to do a bias map correction when we have ignored resolution.

## I. 1 Simple one dimensional example of need for bias map correction

To understand why calibration does not fully remove resolution-induced bias and why we need to apply an after-the-fact "Bias Map Correction" to get back to the correct values after doing the fit to data, we begin by rederiving the moments efficiency calibration method of Section 6.2 in notation more suited to explicating this issue.

The full resolution smeared PDF is given by:

$$
\begin{equation*}
\tilde{g}\left(y^{\prime} ; a\right)=\int d y \mathcal{F}\left(y^{\prime} ; y\right) g(y ; a) \tag{I.1}
\end{equation*}
$$

where a slightly modified version of the notation of Section 6.2 has been used: $g$ is the true, unsmeared PDF (this is $g^{\text {th }}$ in Section 6.2), $\tilde{g}$ is the smeared PDF (called $g^{\text {obs }}$ (after normalization) before), $y^{\prime}$ and $y$ are as before the smeared and true (unsmeared) kinematic variables, and $\mathcal{F}$ is the detector response function (same as before), and $a$ has been used in place of $\mu$ to indicate the parameters. We also assume a multi-parameter PDF implicitly as before, but use $a$ to indicate all of these parameters.

In order to make a cleanly factorizable likelihood, we use the following approach. Rewrite eq.(I.1) as follows:

$$
\begin{equation*}
\tilde{g}\left(y^{\prime} ; a\right)=\int d y \mathcal{F}\left(y^{\prime} ; y\right) g\left(y ; a_{0}\right) \times \frac{g(y ; a)}{g\left(y ; a_{0}\right)} \tag{I.2}
\end{equation*}
$$

where $a_{0}$ is the parameter used to generate some MC sample.
Rewrite the smeared PDF as:

$$
\begin{equation*}
\tilde{g}\left(y^{\prime} ; a\right)=\frac{g\left(y^{\prime} ; a\right)}{g\left(y^{\prime} ; a_{0}\right)} \tilde{g}\left(y^{\prime} ; a_{0}\right) \tag{I.3}
\end{equation*}
$$

with the normalization

$$
\begin{equation*}
\tilde{N}(a)=\int d y^{\prime} \tilde{g}\left(y^{\prime} ; a\right) \tag{I.4}
\end{equation*}
$$

As before, the normalization is essential and is what gives the acceptance correction in this approach. Without it we would not be accounting for the efficiency or resolution, but would just be doing a naive fit to the unsmeared PDF.

To estimate the normalization we can use MC integration by summing over our MC sample generated with $a_{0}$ as follows:

$$
\begin{equation*}
\widetilde{N}(a) \approx \frac{1}{N_{M C}} \sum_{m} \frac{g\left(y_{m}^{\prime} ; a\right)}{g\left(y_{m}^{\prime} ; a_{0}\right)} \tag{I.5}
\end{equation*}
$$

Taking the natural $\log$ of eq.(I.1), our likelihood then becomes (dropping $a$ independent terms):

$$
\begin{equation*}
\ln L(a) \approx \sum_{d} \ln \left(g\left(y_{d}^{\prime} ; a\right)-N_{d a t a} \ln \sum_{m} \frac{g\left(y_{m}^{\prime} ; a\right)}{g\left(y_{m}^{\prime} ; a_{0}\right)}\right. \tag{I.6}
\end{equation*}
$$

where the $d$ sum is over data and the $m$ sum is over MC generated with parameter $a_{0}$.

Now onsider the following PDF with single kinematic variable $\chi$ and single parameter $a$ :

$$
\begin{equation*}
f(\chi ; a)=\frac{1}{\pi}(1+a \cos 2 \chi) \tag{I.7}
\end{equation*}
$$

with $\chi \in\{0, \pi\}$.
To simplify our example, we use a moments method rather than full likelihood to estimate $a$ from a sample of $N$ data points $\left(\chi_{i}\right)$ drawn from the $\chi$-distribution of eq.(I.7).

To apply this we multiply through by $\cos (2 \chi)$ on both sides, integrate over the whole $\chi$ range and divide through by the $\chi$ interval. This yields:

$$
\begin{equation*}
\int d \chi f(\chi ; a) \cos 2 \chi=\frac{1}{\pi}\left(a \int d \chi \cos ^{2} 2 \chi\right) \tag{I.8}
\end{equation*}
$$

Since $\int \cos 2 \chi=0$.
Doing the integral on the right yields:

$$
\begin{equation*}
\int d \chi f(\chi ; a) \cos 2 \chi=\frac{1}{\pi}\left(a \frac{\pi}{2}\right)=\frac{a}{2} \tag{I.9}
\end{equation*}
$$

or

$$
\begin{equation*}
a=2 \int d \chi f(\chi ; a) \cos 2 \chi \tag{I.10}
\end{equation*}
$$

and now as before in Section 6.2.3 we estimate the integral by an MC sum:

$$
\begin{equation*}
\int d \chi f(\chi ; a) \cos 2 \chi \approx \sum_{M C} \frac{\cos 2 \chi}{N_{M C}} \tag{I.11}
\end{equation*}
$$

Thus our estimator $\tilde{a}$ for $a$ is (dropping the " MC " so it is henceforth implicit)

$$
\begin{equation*}
\tilde{a}=\frac{2}{N} \sum \cos \left(2 \chi_{i}\right) \tag{I.12}
\end{equation*}
$$

It is unbiased.

Now suppose there are measurement errors ${ }^{1}\left(\Delta \chi_{i}\right)$ for each $\chi_{i}$ such that we observe only the measured value $\chi_{i}^{\prime}=\chi_{i}+\Delta \chi_{i}$ and that we estimate $a$ naively ignoring the existence of measurement errors. We then obtain

$$
\begin{equation*}
\tilde{a}^{\prime}=\frac{2}{\pi} \sum \cos \left(2 \chi_{i}^{\prime}\right) \tag{I.13}
\end{equation*}
$$

which will differ from the unbiased estimator $\tilde{a}$ - i.e., we will obtain a biased result.

Using the biased estimator $\tilde{a}^{\prime}$ a correction will be needed. The correction can be obtained as follows:

$$
\begin{gather*}
\tilde{a}^{\prime}=\frac{2}{\pi} \sum \cos 2\left(\chi_{i}+\Delta \chi_{i}\right)  \tag{I.14}\\
=\frac{2}{\pi} \sum\left(\cos \left(2 \chi_{i}\right) \cos \left(2 \Delta \chi_{i}\right)-\sin \left(2 \chi_{i}\right) \sin \left(2 \Delta \chi_{i}\right)\right) \tag{I.15}
\end{gather*}
$$

If we average over an infinite ensemble of possible measurement errors, the sine term will average to zero $\left(<\sin 2 \Delta \chi_{i}>=0\right)$. If we assume the resolution is independent of $\chi$ (which is not always the case, but can be taken to be so safely for this simple example), we obtain:

$$
\begin{equation*}
\tilde{a}^{\prime}=\tilde{a} \times\langle\cos 2 \Delta \chi> \tag{I.16}
\end{equation*}
$$

[^6]so the correction factor is the inverse of the mean of the cosine of the errors ( $f_{\text {corr }} \equiv<$ $\cos 2 \Delta \chi>)$. For small errors this is given by
\[

$$
\begin{equation*}
f_{\text {corr }}=<\cos \Delta \chi>\approx 1-2<(\Delta \chi)^{2}=1-2 \sigma_{\chi}^{2} \tag{I.17}
\end{equation*}
$$

\]

If $\sigma_{\chi}=0.2$ (about what it is for Y -frame or diamond frame) then the correction factor is $\mathrm{f}_{\text {corr }}=0.92$. Note that $f_{\text {corr }}=\frac{d \tilde{a}^{\prime}}{d \tilde{a}}$.

The error must, of course, also be corrected as well. In this simple case it is just a factor of $1 / f_{\text {corr }}$ - i.e., we have

$$
\begin{equation*}
\sigma_{\tilde{a}}=\frac{\sigma_{\tilde{a}^{\prime}}}{f_{\text {corr } r}} \tag{I.18}
\end{equation*}
$$

In the multi-parameter case in which we are actually working, we need the full derivative matrix to correct the error, and we see this in Section 8.1.3.

## I. 2 Extra example of need for bias map correction

Now take:

$$
\begin{equation*}
g(y ; a)=C(1+a h(y)) \tag{I.19}
\end{equation*}
$$

where $C$ is a constant and $\int d y h(y)=0$ (i.e., we may assume $h$ is an antisymmetric function for the purposes of this example). The $\cos 2 \chi$ example above is this
type of PDF, as is the PDF in a time-dependent asymmetry analysis. This type of PDF preserves all the essential features of the efficiency calibration method while making the algebra of working it out explicitly somewhat clearer.

Substituting eq.(I.19) into eq.(I.6) and again only keeping parameter-dependent terms yields:

$$
\begin{equation*}
\ln L(a) \approx \sum_{d} \ln \left(1+a h\left(y_{d}^{\prime}\right)\right)-N_{d a t a} \ln \sum_{m} \frac{1+a h\left(y_{m}^{\prime}\right)}{1+a_{0} h\left(y_{m}^{\prime}\right)} \tag{I.20}
\end{equation*}
$$

We maximize the likelihood by solving for $a$ such that

$$
\begin{equation*}
0=\frac{d \ln L}{d a}=\sum_{d} \frac{h\left(y_{d}^{\prime}\right)}{1+a h\left(y_{d}^{\prime}\right)}-N_{d a t a} \frac{1}{\sum_{m} \frac{1+a h\left(y_{m}^{\prime}\right)}{1+a_{0} h\left(y_{m}^{\prime}\right)}} \sum \frac{h\left(y_{m}^{\prime}\right)}{1+a_{0} h\left(y^{\prime}\right)_{m}} \tag{I.21}
\end{equation*}
$$

To solve this, we assume in this example that $a$ and $a_{0}$ are small and expand to first order. We find:

$$
\begin{equation*}
0=\sum h_{d}\left(1-a h_{d}\right)-\frac{N_{d a t a}}{N_{m c}}\left(1-\frac{\left(a-a_{0}\right) \sum h_{m}}{N_{m c}}\right) \sum h_{m}\left(1-a_{0} h_{m}\right) \tag{I.22}
\end{equation*}
$$

where we have defined $h_{i} \equiv h\left(y_{i}^{\prime}\right)$ to compress the notation slightly.
The result of a naive, resolution-uncorrected fit to the MC is:

$$
\begin{equation*}
\tilde{a}_{0}=\frac{\sum h_{m}}{\sum h_{m}^{2}} \tag{I.23}
\end{equation*}
$$

Subsituting this into eq.(I.22) to eliminate $\sum h_{m}$ yields:

$$
0=\sum\left(h_{d}-a h_{d}^{2}\right)-\frac{N_{\text {data }}}{N_{m c}}\left(1-\frac{\left(a-a_{0}\right) \tilde{a}_{0} \sum h_{m}^{2}}{N_{m c}}\right)\left(\tilde{a}_{0}-a_{0}\right) \sum h_{m}^{2}
$$

or rewriting in terms of means (e.g. $\left\langle g>_{d} \equiv \sum_{d} g_{d} / N_{\text {data }}\right.$ ), dividing through by $N_{d a t a}$ and neglecting terms of order $a^{2}$ we find:

$$
\begin{equation*}
0=<g>_{d}-a<h^{2}>_{d}-\left(\tilde{a}_{0}-a_{0}\right)<h^{2}>_{m} \tag{I.24}
\end{equation*}
$$

where a $d$ indicates an average over data and an $m$ indicates average over MC.
Solving for the value of $a$ that solves this equation (call it $\tilde{a}$ ) gives:

$$
\begin{equation*}
\tilde{a}=\frac{<h>_{d}}{\left\langle h^{2}>_{d}\right.}-\left(\tilde{a}_{0}-a_{0}\right) \frac{\left\langle h^{2}>_{m}\right.}{\left\langle h^{2}>_{d}\right.}=\frac{<h>_{d}}{\left\langle h^{2}>_{d}\right.}-\left(\tilde{a}_{0}-a_{0}\right) \tag{I.25}
\end{equation*}
$$

where we have used the fact that for the form of the PDF we are using $\left\langle h^{2}\right\rangle$ does not depend on $a$ and thus does not depend on whether we are using MC or data to estimate average, i.e., $\left\langle h^{2}>_{d} \approx<h^{2}>_{m}\right.$.

This result tells us that if the true value of $a$ is $a_{0}$ then apart from statistical fluctuations we recover the true value.

Suppose that there is no inefficiency and that the only bias is due to resolution. In this case using the naive fit results in a multiplicative bias which we will call $D$ (for "dilution"). We then have for the original true value of $a, a_{t}$ :

$$
\begin{equation*}
\tilde{a}_{t}=\frac{\left\langle h>_{d}\right.}{\left\langle h^{2}>_{d}\right.}=D a_{t} \tag{I.26}
\end{equation*}
$$

and

$$
\begin{equation*}
\tilde{a}_{0}=\frac{\left\langle h>_{m}\right.}{\left\langle h^{2}>_{m}\right.}=D a_{0} \tag{I.27}
\end{equation*}
$$

Substituting this into eq.(I.25) gives us:

$$
\begin{equation*}
\tilde{a}=D a_{t}-(D-1) a_{0} \tag{I.28}
\end{equation*}
$$

So when $a_{t}$ is not $a_{0}$ we do not recover $a_{t}$. The result is biased. The bias $\Delta a_{t}$ is given by

$$
\begin{equation*}
\Delta a_{t}=\tilde{a}-a=(D-1)\left(a_{t}-a_{0}\right) \tag{I.29}
\end{equation*}
$$

so that if $a_{t}$ is very close to $a_{0}$ the bias will not be very large. (This is, of course, always the case for us when we are fitting MC, since we only have MC generated with the CLEO parameters).

Even if $a_{t}$ is near $a_{0}$ and the bias is small we still have not accounted for the effect on the uncertainty. The error estimate we obtain from our fit is $\sigma_{a}^{2}=-1 / \frac{d^{2} \ln L}{d a^{2}}$. In our fit we have

$$
\begin{equation*}
\frac{d^{2} \ln L}{d a^{2}}=N_{d a t a}\left(<h^{2}>_{d}+\tilde{a}_{0}\left(\tilde{a}_{0}-a_{0}\right)<h^{2}>_{d}^{2}\right) \tag{I.30}
\end{equation*}
$$

The only extra term not included here is of order $a^{2}$ and should be neglected to be consistent.

Consider the case of extremely bad resolution, such that $D=0$. In this case there is no information in the measurement and $\frac{d^{2} \ln L}{d a^{2}}$ should be zero indicating an infinite
error. Equation (I.30) does not go to zero as $D \rightarrow 0$, so it cannot be the correct way to estimate the error.

This indicates that we must use the bias map procedure to correct for bias that may be left even after using our calibration procedure, and more importantly to correct the statistical error.

This example has been worked out for a fairly simple case, but illustrates the same basic result we would achieve with a more general PDF.

## Appendix J

## Extra Goodness-of-Fit Validations

## J. 1 GOF on EvtGen samples

In Table J. 1 we show initial tests for comparing $\frac{\chi^{2}}{\text { dof }}$ values between different EvtGen samples, as a control. We see that between statistically independent samples generated with the same parameters, $\frac{\chi^{2}}{\text { dof }}$ is generally quite close to unity. On the contrary, as soon as we compare samples generated with different parameters, $\frac{\chi^{2}}{\text { dof }}$ rapidly increases from unity, and the farther the parameters are apart, the farther it diverges. This behavior is consistent on files with or without cuts, giving us confidence we can proceed to a fully simulated MC sample with all acceptance included.

| Sample 1 | Sample 2 | $\frac{\chi^{2}}{\text { dof }}$ |
| :--- | :--- | :--- |
| Rlll.0 | R111.2 | 0.981 |
| R111.0 | R111.6 | 1.007 |
| R111.2 | R111.6 | 1.013 |
| R111.0 | R111.6 | 1.018 |
| R111.2' | R111.4 | 0.953 |
| Raaa | Rbbb | 2.565 |
| Raaa | Rcce | 8.090 |
| Raaa | Rddd | 17.462 |
| Raaa | Reee | 28.843 |
| Raaa | Rlll'" | 4.886 |

Table J.1: The $\frac{\chi^{2}}{\text { dof }}$ for EvtGen samples, the first two columns are the samples compared, and the third is the $\frac{\chi^{2}}{\text { dof }}$. Sample description: Rlll.0, 2,6 are 112k event statistically independent samples generated with the CLEO parameters (as per the Standard Form Factor Parameter Notation of Section 7.2.1), and with cuts applied. Rlll.2', $4^{\prime}$ are 100k event statistically independent samples generated with the CLEO parameters and no cuts. Raaa,bbb,ccc,ddd,eee,lll" are 100k event samples generated with various sets of parameters. See text for more details.

| Sample 1 | Sample 2 | $\frac{\chi^{2}}{\text { dof }}$ |
| :--- | :--- | :--- |
| SP4A.PK | SP4B.PK | 0.963 |
| SP4A.PK | SP4B.SB | 9.885 |
| SP4A.PK | SP4B.PKBKGD | 13.246 |
| SP4A.SIG | SP4B.SIG | 0.984 |
| SP4A.PKBKGD | SP4B.PKBKGD | 0.996 |
| SP4A.SB | SP4B.SB | 0.934 |

Table J.2: The $\frac{\chi^{2}}{\text { dof }}$ for SP4 samples, the first two columns are the samples compared, and the third is the $\frac{\chi^{2}}{\text { dof. Naming scheme for samples: SP4B.PK is the signal region piece of }}$ SP4B (which has both signal and residual background), SP4B.SIG is the MC-truth selected signal in the signal region piece of SP4B, SP4B.PKBKGD is the MC-truth selected peaking background piece of SP4B, and SP4B.SB is the sideband region sample from SP4B, and the same nomenclature holds for SP4A.

## J. 2 GOF on SP4 samples

In Table J. 2 we do exactly this, first comparing two statistically independent samples generated with the same parameters and seeing again a $\frac{\chi^{2}}{\text { dof }}$ of close to unity note that in this case we are comparing at once both the simulations of the signal and the background since we are only taking the signal region cuts, which intrinsically include the approximately $15-25 \%$ background that remains as a residual in the signal region after all final cuts are made. We next compare the signal region distribution with two types of background: that in the sideband region (which is almost all combinatorial), and that in the peaking region (about $25 \%$ peaking, indicating mostly a correctly reconstructed $D^{*}$ with a wrongly reconstructed lepton for various reasons). In this case we must scale the number of events to be the same (or weight each event in the smaller sample accordingly), and we then find that the $\frac{\chi^{2}}{\text { dof }}$ has increased far above unity.

## J. 3 GOF on Data vs. MC background control samples

Finally, to get a picture of how well our data background compares with our MC , so that we may get a feeling for how valid the final test on data will be, we compared several background control samples of data vs. MC. These initial tests were done on Rel. 8 data and SP3 MC, since that is what the background control sample plots of Section K. 6 were made with.

For Figure K.26, which shows the $\cos \theta_{B Y}$ sideband region (and is mostly populated by charged $B$ decays to $D^{* * ' s}$, which then decay to $D^{*} \mathrm{~s}$ and an extra pion), with about 4 k MC events, we find: $\frac{\chi^{2}}{\frac{1}{d o f}}=0.76$

For Figure K.27, which shows the fake lepton background (where all signal cuts are made except the lepton is chosen to fail the lepton PID cuts), for about 9 k MC events, we find: $\frac{\chi^{2}}{\text { dof }}=1.25$.

For Figure K.28, which shows the uncorrelated lepton background (i.e., in which the $D^{*}$ and lepton came from two separate $B^{0}$ 's), with only about 500 MC events, we find: $\frac{\chi^{2}}{\text { dof }}=1.081$.

For Figure K.29, which shows the combinatorial sideband region background, with about 5 k MC evts, we find: $\frac{\chi^{2}}{\text { dof }}=0.787$.

See Table J. 3 for a summary of these results.
Since interpretation of GOF when there are many zero bins or bins with few events is not fully straightforward, the general lesson here is that as long as the $\frac{\chi^{2}}{\mathrm{dof}}$ is less than about 1.3 or so, we are in a range where the match between two distributions

| Sample | $\frac{\chi^{2}}{\text { dot }}$ |
| :--- | :--- |
| $\cos \theta_{B Y}$ sideband | 0.76 |
| Fake Lepton | 1.25 |
| Uncorrelated Lepton | 1.081 |
| combinatorial sideband region | 0.787 |

Table J.3: The $\frac{\chi^{2}}{\text { dof }}$ for background control samples in data vs. MC. See text in Section J. 3 for more details.
(reweighted or not) is good, but when $\frac{\chi^{2}}{\text { dof }}$ goes significantly greater than this value, the distributions are definitely quite different.

## Appendix K

## Extra MC vs. Data Plots

Text descriptions of the plots in this Appendix (not necessary for the central analysis thrust, but interesting to peruse) are given in Section 7.2 (and in the captions here also).

## K. 1 Signal and Sideband Region SP4 vs. Data Plots By Lepton

## Type



Figure K.1: Kinematic Variables in the signal window for electrons only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.


Figure K.2: Kinematic Variables in the sideband region for electrons only, these indicate the amount of agreement in the combinatorial background between Data and MC.


Figure K.3: Kinematic Variables in the signal window for muons only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.


Figure K.4: Kinematic Variables in the sideband region for muons only, these indicate the amount of agreement in the combinatorial background between Data and MC.

## K. 2 Signal and Sideband Region SP4 vs. Data Plots By $D^{0}$ Mode



Figure K.5: Kinematic Variables in the signal window for the $D^{0} \rightarrow K^{-} \pi^{+}$mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.


Figure K.6: Kinematic Variables in the sideband region for the $D^{0} \rightarrow K^{-} \pi^{+}$mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.


Figure K.7: Kinematic Variables in the signal window for the $D^{0} \rightarrow K^{-} \pi^{+} \pi^{+} \pi^{-}$mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.


Figure K.8: Kinematic Variables in the sideband region for the $D^{0} \rightarrow K^{-} \pi^{+} \pi^{+} \pi^{-}$mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.


Figure K.9: Kinematic Variables in the signal window for the $D^{0} \rightarrow K \pi \pi^{0}$ mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.


Figure K.10: Kinematic Variables in the sideband region for the $D^{0} \rightarrow K \pi \pi^{0}$ mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.


Figure K.11: Kinematic Variables in the signal window for the $D^{0} \rightarrow K_{S} \pi \pi$ mode only, these indicate decent agreement between both the true signal and backgrounds after all final cuts are applied.


Figure K.12: Kinematic Variables in the sideband region for the $D^{0} \rightarrow K_{S} \pi \pi$ mode only, these indicate the amount of agreement in the combinatorial background between Data and MC.

## K. 3 Angular Kinematic Variables with cuts on other kinematic

 variables

Figure K.13: The distributions of $\cos \theta_{V}$ and $\cos \theta_{\ell}$ for different regions of w in the signal window for both Data and MC after the final cuts are applied. Left-hand column is $\cos \theta_{V}$, right-hand column $\cos \theta_{\ell}$. Top row: upper half of $w$ range, bottom row,lower half of $w$ range.


Figure K.14: The distribution of $\chi$ for different regions of $\cos \theta_{V}$ in the signal window for both Data and MC after the final cuts are applied. Left column top row is lowest quarter of $\cos \theta_{V}$ range, right column top row is the middle two quarters of $\cos \theta_{V}$ range and left column bottom row is the highest quarter of $\cos \theta_{V}$ range. Note that these distributions are also visible in the 2 dimensional correlation plots for $\cos \theta_{V}$ and $\chi$. Carefully studying the PDF will show that these plots show behavior concordant with it.

## K. 4 SP4 vs. Data 2D Kinematic Variable Correlation Plots



Figure K.15: Two dimensional correlation plots, MC are on the left, data on the right for the kinematic variables in the signal window. The top three pairs are between w and the three angles, the bottom three pairs between the three angles themselves. We see a good agreement between Data and MC here (since we do not overlay the plots, these and all following 2D plots are not normalized to the same number of events).


Figure K.16: Two dimensional correlation plots, MC are on the left, data on the right for the kinematic variables in the high sideband region. The top three pairs are between w and the three angles, the bottom three pairs between the three angles themselves. We see a good agreement between Data and MC here, indicating good combinatorial background modeling in the MC.


Figure K.17: The 2D correlation plots between the Kinematic Variables, shown is the Data histogram divided by the MC histogram for the signal window. The numbers given are the bin entries.


Figure K.18: The 2D correlation plots between the Kinematic Variables, shown is the $\chi 2$ contribution per bin for the Data histogram divided by the MC histogram for the signal window.


Figure K.19: Slow pion $p_{T}$ for EvtGen produced files, the legend is as described in Section 7.2.1. Left are the XLL plots, Right are LXL.


Figure K.20: These are the LLX plots for slow pion $p_{T}$ for EvtGen files

## K. 5 Slow pion momentum in SP4 vs. Data and EvtGen




Figure K.21: Slow pion momentum in the signal window, indicates our MC is doing a decent job at simulating the signal slow pion momentum distribution (with some of the variation coming from the form factor differences in data and MC).


Figure K.22: Slow pion momentum in the sideband region, indicates our MC is doing well at simulating the background slow pion momentum distribution.


Figure K.23: Lepton lab momentum for EvtGen produced files, the legend is as shown and described in Section 7.2.1. Left are the XLL plots, Right are LXL.


Figure K.24: These are the LLX.

## K. 6 Data/MC Control Plots



Figure K.25: Comparison of kinematic variables of signal-type events in data and MC (all signal cuts as listed in Ch. 5).


Figure K.26: Comparison of kinematic variables of $B^{ \pm}$-type background in data and MC (signal cuts except for $\cos \theta_{B Y}$ outside of the $[-1 \rightarrow 1]$ window).


Figure K.27: Comparison of kinematic variables of fake lepton-type background in data and MC (all signal cuts except for non-VT electron or muon selected at final step).


Figure K.28: Comparison of kinematic variables of uncorrelated lepton ( $D^{*}$ and lepton from different B's) type background in data and MC (signal cuts except for $\cos \theta_{D^{*} \ell}>0$ selected, so $D^{*}$ and lepton point into same hemisphere, which is strongly disfavored in true $D^{*} \ell \nu$ decay).


Figure K.29: Comparison of kinematic variables of combinatoric sideband region type background in data and MC (signal cuts except for high mass sideband region $0.165>$ $\delta m>0.150 \mathrm{GeV}$ selected).

## Appendix L

## EvtGen EMSE and DUBS errors

## validation

## L. 1 EMSE error - Analytic EvtGen Results

As this is a fairly involved computation, it is desirable to check the results on EvtGen samples where we can isolate this effect.

We take for this purpose two EvtGen samples (A and B) generated with 200k events each, and then apply the standard three canonical cuts listed in Appendix D. These reduce both the samples to about 111 k events each. We now take sample A and compute the calibration function (i.e., obtaining the $18 \xi$ acceptance calibration constants from the sample as per Section 6.2.3 ). We then fit to Sample B using this calibration function, and extract the error matrix from this fit, the results are these:


We then evaluate the sums (8.28) to (8.30) by looping over all accepted events of Sample A, substituting the results into eq. (8.25). Finally we substitute both eq. 8.25 and the above error matrix into eq. (8.17), evaluating for all 9 pairs of indices $\alpha, \alpha^{\prime}$, and we obtain:

$$
\mathbf{e}_{\alpha \alpha^{\prime}}^{E M S E}=\left\langle\delta \mu_{\alpha} \delta \mu_{\alpha^{\prime}}\right\rangle=\left(\begin{array}{ccc}
0.000184 & -0.000112 & 0.000085  \tag{L.1}\\
-0.000112 & 0.000186 & -0.000119 \\
0.000085 & -0.000119 & 0.000098
\end{array}\right)
$$

Taking the square root of the diagonal terms results in:

$$
\begin{align*}
& \sigma_{R_{1}}^{\mathrm{EMSE}}=\sqrt{\left(\mathbf{e}_{11}^{E M S E}\right)}=0.013567  \tag{L.2}\\
& \sigma_{R_{2}}^{\mathrm{EMSE}}=\sqrt{\left(\mathbf{e}_{22}^{E M S E}\right)}=0.013651  \tag{L.3}\\
& \sigma_{\rho^{2}}^{\mathrm{EMSE}}=\sqrt{\left(\mathbf{e}_{33}^{\mathrm{EMSE}}\right)}=0.009922 \tag{L.4}
\end{align*}
$$

or an EMSE error contribution of about ( $.014, .014, .010$ ).

## L.1.1 EMSE error - Comparison of EvtGen Results to ToyMC

To get a feeling for this without any contribution from the analytic calculation, we may generate multiple EvtGen samples with the CLEO parameters and again take the canonical cuts (these are of the same type as in the previous section, each about 111 k events). We did this for 45 samples, and then fit repeatedly to the same sample, and plotted a distribution of the results. The central value is expected to be off by about the Minuit error of fitting to a single sample even with infinite calibration statistics but the RMS width of these distributions is expected to give a rough estimate of the EMSE error (at least for EvtGen samples, which is what we are working with here to be free of smearing and all other possibly convoluting effects). We can see the distributions in Figure L.1.

As can be seen from the Figure L. 1 the RMS widths are (.015, . $013, .010$ ), which are quite close to the analytic results calculated above, giving us confidence that our analytic method is sound, and we are able to use it for the full MC (SP4) estimates which are the important results, since we use a SP4 Signal MC sample to do our acceptance correction calibration for the ultimate result on the data.

## L.1.2 EMSE error - Comparison of Analytic Calibration Shift Error to Calcu-

## lated Error

Another check we may make on the error estimate is to use a technique for evaluating it described in [27] (Chapter 4). We define four quantities in this technique:


Figure L.1: Fit distribution using 90 EvtGen calibration files, fitting to the same sample.

- $x_{M}(M)=$ The parameters resulting from a calibration on MC and fit to the MC (i.e., a self-fit, which must result in exactly the input parameters by the self-calibration theorem of Section 6.2.5)
- $x_{D}(M)=$ The parameters resulting from a calibration on MC and fit to the Data
- $x_{M}(0)=$ The parameters resulting from no calibration (i.e., only normalizing by the analytic function, not accounting for acceptance at all) and fit to MC
- $x_{D}(0)=$ The parameters resulting from no calibration (i.e., only normalizing by the analytic function) and fit to Data

We then define two difference functions:

$$
\begin{gather*}
s_{D}=x_{D}(M)-x_{D}(0)  \tag{L.5}\\
s_{M}=x_{M}(M)-x_{M}(0) \tag{L.6}
\end{gather*}
$$

Up to quantities of order $\mathcal{O}\left(x_{M}-x_{D}\right)$ we expect these differences to be the same because the calibration is being done on the same MC Set, and the only variation between the two is the statistical one in the data set. Further, since indeed $x_{M}(M)=x_{M}, s_{D}=s_{M}$ implies that $s_{D}=x_{M}-x_{M}(0)$, and since $x_{M}$ is a fixed input, the error matrix in obtaining $s_{D}$ should be that in obtaining $x_{M}(0)$.

The central issue here is that pointed out above - that it does not take the statistical variation of the Data account into the equivalence, so indeed, denoting by " e " the error matrix, we expect:

$$
\begin{equation*}
\mathbf{e}_{\mathrm{D}}=\mathbf{e}_{\mathrm{M}}+\mathbf{e}_{\text {stat'l data }} \tag{L.7}
\end{equation*}
$$

Without attempting to quantify this difference, we may simply do the fit of $x_{M}(0)$ to see how the error matrices compare, and what we find is:

$$
\mathbf{e}_{\mathrm{M}}=\left(\begin{array}{ccc}
0.000135 & -0.000058 & 0.000043  \tag{L.8}\\
-0.000058 & 0.000110 & -0.000079 \\
& & \\
0.000043 & -0.000079 & 0.000097
\end{array}\right)
$$

When we compare the diagonal terms here versus those of the correctly computed error matrix $\mathbf{e}_{\mathbf{D}}$ which they correspond to, eq.(L.1), we find that they are of the correct order, but low by an average of about $20 \%$ (similarly for the off-diagonal terms, which are also correctly signed relative to the full matrix). This gives us an estimate than of the extra error that is not being taken into account by the above rough method of estimation, and is another confidence cross check telling us that we indeed have an accurate method for obtaining this error.

## L. 2 DUBS error - Analytic Results on Evtgen Samples

As with the EMSE validation, it is desirable to check these analytic formulas results on Evtgen samples where we can isolate this effect.

We take for this purpose one Evtgen sample (A) of 100k events, generated with CLEO signal parameters, and two Evtgen samples (B and C) of 20k events each generated with the "background parameters" ( $R_{1}=0.33, R_{2}=1.26, \rho^{2}=1.13$ ) - these are the parameters that Minuit gives us if we require it to fit the background with our signal PDF.

Note that in no way are we claiming that this is a good fit or that the signal PDF represents the shape of the background in four dimensions well - we simply wanted to generate event samples with parameters significantly different from the signal parameters so we had a large separation and no overlap problems, and this was an easy method to do that. In fact, it is easily seen that fitting the different categories of background results in fit parameters that are very different from each other, so that simply saying "the background parameters are these" makes little sense, since clearly the various types of background are populating completely different regions of the kinematic phase space, and the "background parameters" we have given are just some rough average of all these which is what Minuit gives us when forced to fit the set of background events with the only PDF we have readily available, that of the signal.

In fact, for this study, as we are adding and subtracting the set of toy "background events", the parameters do not matter, as long as they're the same for the two distributions being added and subtracted.

We now add Event Sample B to Sample A, then subtract Sample C, calibrate with the analytic normalization function, run the fit and extract the error matrix, the results are these:

| Parameter |  |  | Fit | Result | Minuit Error |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | R1 | 1.1 | $8617 e+00$ | $1.10341 \mathrm{e}-02$ |
|  | 2 | R2 | 7.2 | 3395e-01 | $1.26869 \mathrm{e}-02$ |
|  | 3 | rho2 | 9.0 | 7128e-01 | $1.09770 \mathrm{e}-02$ |
| ERROR MATRIX. |  |  |  |  |  |
|  | 1 | 2 |  | 3 |  |
| $11.218 e-04-5.312 e-05 \quad 3.633 e-05$ |  |  |  |  |  |
| $2-5.312 e-05 \quad 1.610 e-04-1.108 e-04$ |  |  |  |  |  |
| 3 | $3.633 \mathrm{e}-05-1.108 \mathrm{e}-04 \quad 1.205 \mathrm{e}-04$ |  |  |  |  |

We then evaluate the sums of eq.(8.41) by looping over all events in Sample C. Finally we substitute the above error matrix as well into eq.(8.41), evaluating for all 9 pairs of indices $\alpha, \alpha^{\prime}$, and we obtain:

$$
\mathbf{e}_{\alpha \alpha^{\prime}}^{D U B S}=\left(\begin{array}{ccc}
0.000080 & -0.000044 & 0.000031  \tag{L.9}\\
-0.000044 & 0.000072 & -0.000052 \\
0.000031 & -0.000052 & 0.000052
\end{array}\right)
$$

Taking the square root of the diagonal terms results in:

$$
\begin{align*}
& \sigma_{R_{1}}^{\text {DUBS }}=\sqrt{\mathbf{e}_{11}^{D U B S}}=0.008919  \tag{L.10}\\
& \sigma_{R_{2}}^{\text {DUBS }}=\sqrt{\mathbf{e}_{22}^{D U B S}}=0.008494  \tag{L.11}\\
& \sigma_{\rho^{2}}^{\text {DUBS }}=\sqrt{\mathbf{e}_{33}^{\text {DUBS }}=0.007224} \tag{L.12}
\end{align*}
$$

or a DUBS error contribution for these Evtgen samples of about (.009, .008, .007).
Note that in this case we have used $\eta=1$ since the "luminosity" of the data and background control sample is identical (i.e., we added and subtracted exactly 20 k evts).

## L.2.1 DUBS error - Comparison of Analytic Results to Evtgen ToyMC

To get a feeling for this without any contribution from the analytic calculation, we may take the same 100k event Sample A above that was generated with CLEO parameters, then generate many pairs of 20k event Evtgen samples with the "background parameters" add and subtract these background samples, and run the fit repeatedly. We did this for 100 pairs of background samples and plotted a distribution of the results. The central value is expected to be off by about the Minuit statistical error of the "data" Sample A, but the RMS width of these distributions is expected to give an estimate of the DUBS error (at least for Evtgen samples, which is what we are working with here to be free of smearing and all other possibly convoluting effects). We can see the distributions in Figure L.2.

As can be seen from the Figure L. 2 the RMS widths are (.009, . $008, .006$ ), which are quite close to the analytic results calculated above, giving us confidence that our analytic method is sound, and we are able to use it for the SP4 estimates, which are the important results, since again we use a SP4 Generic MC sample to provide the background which we use as our ultimate control sample to be subtracted from the data.


Figure L.2: Fit distribution using 100 pairs of background control files, fitting to the same initial sample generated with CLEO parameters (see text).

## Appendix M

## $M_{D^{*}}-M_{D^{0}}(\delta M)$ Systematic Error

## Details

While it is clear from visual inspection of Fig. 8.2 (in the main text body) that single Gaussian fits do not capture the full structure of the $\delta m$ peak (i.e., two points are obviously left out), and in fact for realistic fits where this is significant, including those for the measurement of the BR of $B^{0} \rightarrow D^{*-} \ell^{+} \nu$ and $\left|V_{c b}\right|$, a double gaussian plus a threshold function is normally used which results in a much better $\chi^{2}$, this fit of a single Gaussian plus threshold is sufficient for our purposes, as we do not anywhere use the shape in our fit, but only make a hard final cut on $\delta m$, and in Section 8.4 .4 we are interested in the difference in this width between Data and MC.

In fact, the double Gaussian fit results in three parameters: two widths, for the inner core and wider Gaussian (which physically take into account the difference in reso-
lution between the slow pion tracks with DCH hits and those with SVT hits only) as well as the fraction between them - and it is simply not clear exactly which of these widths would be the optimal one to use to reduce the $\mathrm{S} / \mathrm{B}$ ratio for our particular analysis (and in fact, because the shape is so significant for them, for the BR and $\left|V_{c b}\right|$ analyses they normally do two separate cuts for these different classes of events, depending on whether the slow pion did or did not have DCH hits). We have thus used the canonical cut determined in older analyses of $2.5 \sigma$, with $\sigma$ taken to be 1 MeV , about a mean of 145 MeV , which we checked by hand does quite well in including the large majority of correct signal, without accepting an undue amount of the slowly rising combinatorial background (in general the shape of $\delta m$ in the peaking backgrounds is not very different from the signal, as can be checked in control samples, so that narrowing or widening the $\delta m$ cut will not result in any gain in $S / B$ with respect to these backgrounds).

## Appendix $\mathbf{N}$

## FSR SystErr Details

The detailed steps to assigning a null result for the FSR error from Evtgen are layed out here.

To get a sense of the overall size, we first took large samples with no acceptance cut and normalized using just the analytic function (as in Sec.C.2). First, a simple fit to a standard 1M event file with no FSR is done to check the statistical size of the error (note there is no EMSE error involved when the analytic function is used for calibration), and the results shown in Table N. 1 (in this section all input parameters will be taken as the CLEO parameters ( $R_{1}=1.1800, R_{2}=0.7200, \rho^{2}=0.9200$ ). The shifts from the inputs here, $(+.0007,+.0013,-.0018)$, are clearly well within the one $\sigma$ error.

We then took the same no-acceptance fit with analytic normalization to a file with FSR turned on, the results are in Table N.2, and the shifts are (-.0063, $+.0033,+.0007$ ). While these are all within two $\sigma$, they were large enough to warrant looking at more

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.1800 | 1.1807 | 0.0035 |
| $R_{2}$ | 0.7200 | 0.7213 | 0.0029 |
| $\rho^{2}$ | 0.9200 | 0.9182 | 0.0035 |

Table N.1: Analytic function calibration on LLL Evtgen sample with FSR off.
closely, so we then took the more realistic case of testing on files with actual acceptance requiring real calibration.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.1800 | 1.1743 | 0.0035 |
| $R_{2}$ | 0.7200 | 0.7233 | 0.0040 |
| $\rho^{2}$ | 0.9200 | 0.9207 | 0.0035 |

Table N.2: Analytic function calibration on LLL Evtgen sample with FSR on.

The first test case we tried was calibration on a no-FSR sample and fitting to another no-FSR sample. We took the size of the calibration sample to be originally 2 M events, which becomes about 1.1 M events after cuts, and the fit sample to be originally 1 M events, about 0.57 M events after the cuts [though these very large fit sample sizes do extend the fit time considerably, we wanted to reduce the size of the statistical errors associated with the fits to a fairly small level in order to isolate the FSR effects; further we take the calibration sample to be even larger to reduce the associated shifts from EMSE here (which don't show up in the statistical error), again to isolate the FSR effects]. The results are in Table N.3, and the shifts are (-.0037, +.0089, -.0069) which has $R_{1}$ falling within one $\sigma$ and are making just slightly larger than the one $\sigma$ the errors for $R_{2}$ and
$\rho^{2}$ (and likely within the error if EMSE was taken into account).

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.1800 | 1.1763 | 0.0063 |
| $R_{2}$ | 0.7200 | 0.7290 | 0.0061 |
| $\rho^{2}$ | 0.9200 | 0.9132 | 0.0049 |

Table N.3: Real calibration on another LLL Evtgen sample (no FSR) with cuts (see text).

The next case we tried was calibration on a no-FSR sample and fitting to a sample with FSR, with similar sample sizes as above (this is the one that more closely corresponds to what we will actually be doing, as in the real case the Data has FSR "automatically on", and we only have the ability to choose whatever we want for the calibration file). The results are in Table N.4, and the shifts are (+.0129, $+.0067,-.0039$ ) - these now seem too large to account for simply statistically.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.1800 | 1.1763 | 0.0063 |
| $R_{2}$ | 0.7200 | 0.7290 | 0.0061 |
| $\rho^{2}$ | 0.9200 | 0.9132 | 0.0049 |

Table N.4: Real calibration on another LLL Evtgen sample (FSR on) with cuts (see text).

We then decided to try calibration on samples with FSR and fitting to a sample with FSR, with similar sample sizes as above. The average results of five fits are in Table N.5, and the statistical widths are as shown in the table, the shifts are: (+.0003, -.0022, -.0025). These are well within the statistical errors shown in the table above (which are similar for these samples because the sizes are the same), and we can now rest and say
that if we use for calibration files that have FSR included in them, there is no additional error to be assigned from this source.

| Parameter | Input Value | Fit Value | Width from five fits |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.1800 | 1.1803 | 0.0036 |
| $R_{2}$ | 0.7200 | 0.7178 | 0.0038 |
| $\rho^{2}$ | 0.9200 | 0.9175 | 0.0041 |

Table N.5: Real calibration on another LLL Evtgen sample (FSR on) with cuts (see text).

However, we might try to still go one step better because we know how: we instituted a procedure of including FSR and yet getting back to lepton 4 -vectors that should be closer to the original ones: we take a cone around the original direction of the lepton, find any possible photons within a given opening angle of the cone (in practice, there are only one (about 20\% of the time) or two (about 1\% of the time) ) and add their 4 -vectors back to that of the lepton. We then re-calculate the kinematic variables of this event, and output these instead into the calibration file. A single test case of this is shown in Table N.6, with shifts (-.0053, $+.0023,-.0054)$ - these are also within the one $\sigma$ error (the last just slightly larger), so this realistic procedure seems also to have ameliorated the problem of fitting with a no-FSR file.

| Parameter | Input Value | Fit Value | Minuit Error |
| :--- | :--- | :--- | :--- |
| $R_{1}$ | 1.1800 | 1.1747 | 0.0064 |
| $R_{2}$ | 0.7200 | 0.7224 | 0.0062 |
| $\rho^{2}$ | 0.9200 | 0.9147 | 0.0050 |

Table N.6: Real resummed calibration on another LLL Evtgen sample with cuts (see text).
(Though because the photons become very soft generally as one moves off-axis from the lepton, so likely the angle of the cone makes little difference, for the record the cone size we took was was $\theta=36$ degrees, corresponding to $\cos (\theta)=0.8$, and about $95 \%$ of the FSR photons in the event do fall within this cone. ).

To do a more statistically meaningful test, we ran 10 samples using this resummming procedure, calibrated in turn from each of these, and fit to the same sample as before (which has FSR on). The result of this are widths of (.0044, .0050, .0041), and the distribution can be seen graphically in Fig. N.1. These widths are also well within the statistical error for this fit, and we may conclude that if we using a resumming procedure on the calibration files, we do not need to assign an error from this source.

However, we cannot argue that this resumming method is intrinsically better than fitting with a FSR-on file from these tests, thus we will stay with the simpler latter procedure and also assign a null result to the error from FSR.


Figure N.1: Distribution of fits with FSR resummed photon calibration files, fit to a file with FSR on - the distributions are not very Gaussian (this is only 10 points), but what is important is the small size of the RMS spread.


[^0]:    ${ }^{1}$ Cosmos, the PBS television series.
    ${ }^{2} 1900$ at the British Association for the Advancement of Science.

[^1]:    ${ }^{3}$ The first several sections of this Chapter are modified versions of Chapter 1 of [5], used by permission and gratefully acknowledged.
    ${ }^{4}$ The birthplace of the World Wide Web.

[^2]:    ${ }^{1}$ After this work was completed, we extended the analysis to include the possibility of non-zero coefficients for the $(w-1),(w-1)^{2}$ terms, and determined the initial $R_{1}, R_{2}$ constant terms. As expected, they do not vary appreciably from the $w$-averaged values, see [46] for details.

[^3]:    ${ }^{1}$ Several primary sections of this Chapter are modified versions of Ref.[5] which heavily overlapped and was done in part in conjunction with this work up as far as Event Reconstruction, and permission to use material from [5] is very gratefully acknowledged to C.M. LeClerc.

[^4]:    ${ }^{2}$ Boom.

[^5]:    ${ }^{1}$ See Ref. [46] for a slightly different way of deriving the same result as this Subsection.

[^6]:    ${ }^{1} \Delta \chi_{i}$ is not the uncertainty (a.k.a. resolution) $\sigma_{\chi}$, but the actual error made on measurement $i$.

