LBL-36353
UC-414

Lawrence Berkeley Laboratory

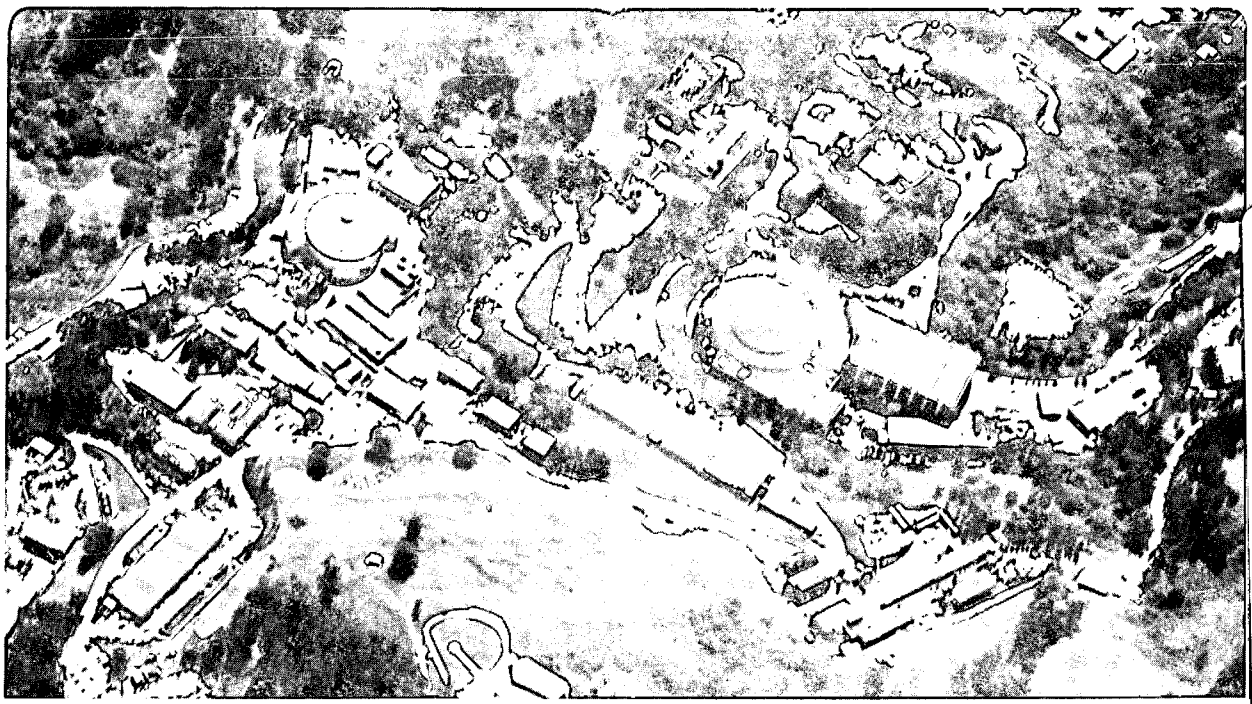
UNIVERSITY OF CALIFORNIA

Physics Division

Study of $b\bar{b}$ Production in e^+e^- Annihilation at $\sqrt{s}$ = 29 GeV with the Aid of Neural Networks

D.J. Lambert
(Ph.D. Thesis)

November 1994



REFERENCE COPY
Does Not
Circulate
Bldg. 50 Library.
LBL-36353
Copy 1

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-36353
UC-414

Study of $b\bar{b}$ Production in e^+e^- Annihilation at $\sqrt{s} = 29$ GeV with the Aid of Neural Networks

David Joel Lambert
Ph.D. Thesis

Department of Physics
University of California

and

Physics Division
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

November 1994

This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

LBL-36353
UC-414

Study of $b\bar{b}$ Production in e^+e^- Annihilation at $\sqrt{s} = 29$ GeV with the Aid of Neural Networks

David Joel Lambert
PhD Thesis

Department of Physics
University of California, Berkeley
and
Lawrence Berkeley Laboratory
University of California, Berkeley

November 15, 1994

Abstract

We present a measurement of $\sigma(b\bar{b})/\sigma(q\bar{q})$ in the annihilation process $e^+e^- \rightarrow q\bar{q} \rightarrow$ *hadrons* at $\sqrt{s} = 29$ GeV. The analysis is based on 66 pb^{-1} of data collected between 1984 and 1986 with the TPC/2γ detector at PEP. To identify bottom events, we use a neural network with inputs that are computed from the 3-momenta of all of the observed charged hadrons in each event. We also present a study of bias in techniques for measuring inclusive $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$ production in the annihilation process $e^+e^- \rightarrow b\bar{b} \rightarrow$ *hadrons* at $\sqrt{s} = 29$ GeV, using a neural network to identify bottom-quark jets. In this study, charged particles are identified by a simultaneous measurement of momentum and ionization energy loss (dE/dx).

This work is supported by the United States Department of Energy under Contract DE-AC03-76SF00098.

Acknowledgements

I consider it a privilege to have participated, with so many gifted people, in a scientific endeavor that aimed to better understand the fundamental workings of the universe. It was a tragedy for all involved that the experiment had to be terminated before we could reap the rewards of the many person-years of labor spent readying it for its high-luminosity running. We can now only dream of what we might have accomplished with such a powerful detector.

I am indebted to many people on the TPC/Two-Gamma experiment who gladly aided me in my quest to do good physics and to earn a doctorate. Most of all, I am indebted to Michael Ronan, who patiently guided me through my entire career as a graduate student on the TPC/Two-Gamma experiment. He provided me with many good ideas that helped make this analysis what it is (e.g. using a neural network), and it was he who encouraged me during my low points. I am very grateful to Ron Ross, who, with Michael Ronan, served as my dissertation advisor; he asked many good questions, always reminded me of the big picture, and always treated me warmly.

I am also grateful to Gerry Lynch, who taught me much of what I now know of statistics, $\mathrm{d}E/\mathrm{d}x$, data processing, and optimization, and who always welcomed my questions. Gerry and Mike also did the work of including the CLEO Monte Carlo into the standard TPC/Two-Gamma Monte Carlo. Phillipe Eberhard came up with the idea of requiring $f_{MC}^{QED} = f_{meas}^{QED}$ in Section 7.6, he is chiefly responsible for creating the binned maximum likelihood fit method described in Section 7.3.2, and his perceptive responses to my analysis progress reports were invaluable. I had many good discussions with Jeremy Lys and Hiro Yamamoto. Lynn Stevenson, Marjorie Shapiro, Orin Dahl, Al Clark, Werner Koellner, Jim Dodge, John Waters, and Tim Edberg all provided valuable assistance. Shigeki Misawa provided me with much information that was valuable for my job-hunt.

Allen Nicol joined the experiment about when I did. I treasured his companionship in the trenches. I shall never forget the gnarly rubber band fights with Brent Corbin. I thank the DØ collaboration at LBL for allowing me access to their workstations. I am grateful to Allen, Tim Edberg, Glen Cowan, and Jack Eastman for letting me 'borrow' their figures. I wish to thank Selig Kaplan, Marjorie Shapiro, Ron Ross, and Mike Ronan, the members of the committee who read this

dissertation.

I am forever indebted to David for his personal guidance. I have been deeply enriched by the friendship of Doyle, Jan, Val, Joyce, and Robin, and the love and friendship of Karen, Tina, and Tracy. Doyle taught me the true meaning of friendship: he typed in a large portion of this dissertation over many hours when I was unable to type because of tendonitis. My parents are world class parents, even when they asked for the latest estimate of my graduation date. AGSE and HAI made life even more interesting.

This dissertation is dedicated to the wonderful people and land of the San Francisco Bay Area.

Contents

Acknowledgements **ii**

1 Overview **1**

2 The Theory of Hadron Production in e^+e^- Annihilation **3**

2.1 Quarks and Their Properties 3

2.2 Production of $q\bar{q}$ in e^+e^- Annihilation 5

2.3 QCD in e^+e^- Annihilation 7

2.3.1 Fixed Order Computations in QCD 7

2.3.2 The Leading Logarithm Approximation of QCD 8

2.3.3 Where Perturbative QCD Fails 8

2.4 Models of Hadronization 9

2.4.1 Independent Fragmentation 9

2.4.2 String Fragmentation 12

2.4.3 Cluster Fragmentation 17

2.4.4 The Peterson Fragmentation Function 19

2.5 The Properties of Heavy Quark Events 21

2.6 Previously Used Heavy Quark Event Tags 22

2.7 Corrections to $\sigma(b\bar{b})/\sigma(q\bar{q})$ 24

3 The TPC/Two-Gamma Experiment **27**

3.1 The TPC/Two-Gamma Detector 27

3.2 The Time Projection Chamber 32

3.3 Calibration of the TPC 34

4 Particle Identification Using the TPC **35**
4.1 The Measurement of Momentum 35
4.2 The Theory of dE/dx Energy Loss 36
4.3 dE/dx Resolution Parameterization 39

5 Event Reconstruction, Selection, and Simulation **44**
5.1 Event Reconstruction 44
5.2 Event Selection 44
5.3 Event Simulation 47
5.3.1 The Event Simulation Software Package 47
5.3.2 Tuning the Peterson Parameterization 48
5.3.3 Tuning the Jetset Event Shape Parameters 49

6 Feed-Forward Neural Networks **60**
6.1 Neural Network Architecture 60
6.2 The Training of a Neural Network 62
6.3 Measuring Network Performance 65
6.4 Previous Uses of Neural Networks in High Energy Physics 67
6.4.1 Classification Using Neural Networks 67
6.4.2 Fitting in a Neural Network Output 67

7 A Measurement of the Bottom Event Production Fraction **72**
7.1 The Choice of Neural Network Inputs and Architecture 72
7.2 Training the Event-Tagging Neural Network 74
7.3 The Method for Fitting the Bottom Event Fraction 80
7.3.1 The Extended Maximum Likelihood Method 82
7.3.2 The Event Fraction Likelihood Function 82
7.4 The Bottom Event Fraction 85
7.4.1 The Fit of the Bottom Event Fraction 85
7.4.2 Correcting for Backgrounds 85
7.4.3 Acceptance and Physics Corrections 88
7.5 The Evaluation of Systematic Errors 89
7.5.1 Systematic Errors in the Simulation of $e^+e^- \to q\bar{q}$ 91
7.5.2 Systematic Errors due to the Detector Simulation 92
7.6 The Monte Carlo Bottom Event Fraction 94

7.7 Discussion of the Results 99
7.7.1 How $\sigma(f^{QED})$ Depends Upon $\sigma(f^{QED}_{meas})$ 99
7.7.2 Comparison to LEP Measurements of $\Gamma(Z^0 \rightarrow b\bar{b})$ 100

8 A Study of Bias in Techniques to Measure Charged Hadron Production in Bottom Quark Jets 103
8.1 Introduction 104
8.2 Track and Event Selections 104
8.3 The Jet-Tagging Neural Network 105
8.4 Techniques for Measuring Charged Hadron Production in Bottom Quark Jets 113
8.5 A Test for Bias in Measurements of Charged Hadron Production in Bottom Quark Jets 120
8.6 An Investigation of the Sources of Bias 121
8.6.1 Using Correlations Between Jets in Each Event to Measure Bias 121
8.6.2 All Possible Sources of Correlations 126
8.6.3 Which Sources of Correlation Contribute 127
8.6.4 The Relative Importance of the Sources of Correlation . . . 135
8.6.5 Comparing Monte Carlo Correlations to Experimental Data Correlations 141
8.7 Summary 143

9 Conclusions 145

Appendix A Remainder of the Proof of the Event Fraction Likelihood Function Optimization Method 147
A.1 The Case some but not all $m_{ij} = 0$ 147
A.1.1 The Subcase $\frac{n_j}{\lambda_j} - 1 < \frac{1}{a_{M'}}$ 147
A.1.2 The Subcase $\frac{1}{a_{M'}} \leq \frac{n_j}{\lambda_j} - 1$ 148
A.2 The Case $m_{ij} = 0$ for all i 148

Appendix B F for Any Number of Classes 149
B.1 K Event Classes with Complete Separation 150
B.2 Computing the F's 150

Bibliography 151

List of Tables

2.1 The properties of quarks. 4
2.2 Some hadrons containing heavy quarks. 6
2.3 Properties of jets of different flavors. 22
2.4 Previous hadronization measurements for different quark flavors. . . 24

5.1 Tuned values of the Lund flavor parameters in Jetset 7.2. 48
5.2 $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$. 49
5.3 ϵ_b and ϵ_c for Jetset 7.2. 49
5.4 Variables used for tuning. 51
5.5 Tune of the Lund parameters in Jetset 7.2. 56

7.1 The values of F for the 7 event-tagging network inputs. 74
7.2 Tune of the Lund event shape parameters in Jetset 7.2. 92
7.3 The systematic errors in the Monte Carlo. 93
7.4 f_{meas}^{QED} for five different values of f_{MC}^{QED}. 95
7.5 The best-fit parabola. 96
7.6 The correlation matrix for the five values of f_{meas}^{QED}. 97
7.7 All measurements of the bottom event fraction using a neural network. 101

8.1 The baseline for the investigation of the sources of correlation. . . . 128
8.2 Correlations from fragmentation physics. 129
8.3 Correlations due to ISR and gluon radiation. 130
8.4 Correlations due to ISR and gluon radiation, axes "iterated". 131
8.5 Correlations from event shape/direction measurements. 133
8.6 Correlations due to detector acceptance. 134
8.7 Correlations from event selections. 136
8.8 Correlations caused by interactions in the detector material. 137

8.9 Correlations relative to "Baseline 2". 139
8.10 Correlations relative to "Baseline 3". 140
8.11 The effect of axis "iterating" on correlations. 141
8.12 Correlations in the Jet-Tagging Network Inputs, for Monte Carlo and for experimental data. 142
8.13 Correlations in the Jet-Tagging Network Output, for Monte Carlo and for experimental data. 143

List of Figures

2.1 The Standard Model fermions. 4
2.2 The independent fragmentation process. 10
2.3 Flux lines for an electric dipole and for color dipole. 13
2.4 The yo-yo model of a meson in string fragmentation. 14
2.5 String fragmentation. 16
2.6 Cluster fragmentation. 18
2.7 The Peterson fragmentation function. 20

3.1 The PEP storage ring. 28
3.2 The TPC/2γ detector, 3-D representation. 29
3.3 The TPC/2γ detector, side view. 30
3.4 The TPC/2γ detector, end view. 31
3.5 A TPC sector. 33
3.6 The Time Projection Chamber wires. 33

4.1 Distribution of $\mathrm{d}E/\mathrm{d}x$ energy loss. 37
4.2 Dependence of $\langle \mathrm{d}E/\mathrm{d}x \rangle$ on $\eta = \beta\gamma$. 38
4.3 $\mathrm{d}E/\mathrm{d}x$ as a function of $\log(p)$. 40
4.4 Scatterplot of N, the number of wire hits, as a function of $|\sin\lambda|$ for pions. 41
4.5 Number of pions as a function of $(R-1)/\sigma(N,\lambda)$. 43

5.1 Discrepancy between Monte Carlo and Experiment 14-18 data without and with the τ-2γ selection for the multiplicity, thrust minor, and thrust major. 55
5.2 Comparison of experimental data to tuned Monte Carlo in the tuning variables. 56
5.3 Comparison of experimental data to tuned Monte Carlo in the tuning variables. 57

5.4 Comparison of experimental data to tuned Monte Carlo in the tuning variables. 58

6.1 A one hidden layer neural network. 61
6.2 The DØ electron-hadron calorimeter neural network. 68
6.3 L3 fit of $\Gamma(b\bar{b})$ with a neural network. 69
6.4 DELPHI neural network output distributions. 71

7.1 Event-Tagging Neural Network Inputs 1-4. 75
7.2 Event-Tagging Neural Network Inputs 5-7. 76
7.3 F as a function of the number of hidden nodes for the Event-Tagging Neural Network. 77
7.4 F as a function of the number of patterns per Event-Tagging Network parameter. 78
7.5 F as a function of epoch number for the Event-Tagging Neural Network. 79
7.6 The Experiment 14-18 event-tagging neural network output distribution. 80
7.7 The Monte Carlo event-tagging neural network output distributions. 81
7.8 The fit of the bottom event fraction. 86
7.9 Bottom event efficiency and purity as a function of neural network output. 87
7.10 Measured b-event fraction as a function of Monte Carlo b-event fraction. 96
7.11 The difference between the measured and Monte Carlo b-event fractions. 98
7.12 Geometrical picture of how the error is magnified. 100

8.1 The *uniterated* coordinate system. 107
8.2 The *iterated* coordinate system. 107
8.3 Jet-Tagging Neural Network inputs 3, 6, 9, and 12. 109
8.4 Jet-Tagging Neural Network inputs 1, 4, 7, and 10. 110
8.5 Jet-Tagging Neural Network inputs 2, 5, 8, and 11. 111
8.6 Jet-Tagging Neural Network inputs 13, 14, and 15. 112
8.7 F as a function of the number of hidden nodes for the Jet Neural Network. 113
8.8 F as a function of epoch number for the Jet-Tagging Neural Network. 114

8.9 Jet-Tagging Neural Network output distribution for Experiments 14-18. 115
8.10 Jet-Tagging Neural Network output distributions for Monte Carlo. . 116
8.11 The fit of the bottom jet fraction. 117
8.12 Bottom jet efficiency and purity as a function of neural network output. 118
8.13 The bin-by-bin confidence levels for the scaled pion cross-section to be independent of the network output. 122
8.14 The bin-by-bin confidence levels for the scaled kaon cross-section to be independent of the network output. 123
8.15 The bin-by-bin confidence levels for the scaled proton cross-section to be independent of the network output. 124
8.16 The sum of three independent distributions is not independent. . . 125

Chapter 1

Overview

In the Standard Model of elementary particle physics, fermionic quarks and leptons interact through the exchange of force-mediating bosons.

1. Photons mediate the electromagnetic interaction, which is described by Quantum Electrodynamics (QED).
2. The $W^{\pm}$ and the Z^0, which are unified with the photon in the Weinberg-Salam model, mediate the weak interaction responsible for radioactive decay.
3. Gluons mediate the strong (color) interaction that binds quarks into hadrons. The theory of the color force is Quantum Chromodynamics, abbreviated QCD.

The Standard Model is very successful: it has survived many tests, made many successful predictions, and does not conflict with any experimental observations.

We measure the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$ in this dissertation, since this ratio is a fundamental prediction of the Standard Model that has never been directly measured near 29 GeV. The analysis is based on 66 pb^{-1} of data collected between 1984 and 1986 with the TPC/2γ detector at PEP.

Past measurements of the properties of bottom events used standard methods for identifying, or tagging, bottom events. These methods collected fairly high-purity bottom event samples, but with low efficiency: most bottom events are excluded from the bottom sample. We take a new approach in order to use the data more efficiently: we use a neural network with inputs that are computed from the 3-momenta of all of the observed charged hadrons in each event.

One of the few outstanding problems of the Standard Model is that, as of now, we have neither the theoretical nor the computational ability to make quantitative predictions with QCD for processes with momentum-transfers of the order of 1

GeV/c or less. One of these low-momentum-transfer processes is hadronization, in which colored quarks and gluons are confined within colorless hadrons.

To point the way towards improving our understanding of hadronization, we also present a study of bias in techniques for measuring inclusive $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$ production in the annihilation process $e^+e^- \rightarrow b\bar{b} \rightarrow$ *hadrons* at $\sqrt{s} = 29$ GeV, using a neural network to identify the contribution of bottom-quark jets.

Chapter 2 contains a review of the theory of hadronization in e^+e^- annihilation, a discussion of the previous measurements of hadronization for events produced by different kinds of quarks, and a review of the theory of e^+e^- annihilation that we use to determine the theoretical value of the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$ to compare to our measurement of this ratio. In Chapter 3, we provide a brief description of that part of the TPC/Two-Gamma experiment which is relevant to this measurement, and in Chapter 4, we describe particle identification using the TPC. In Chapter 5, we describe the processing, selection, and simulation of the $e^+e^- \rightarrow$ *hadrons* data. Chapter 6 contains our description of the use of neural networks. In Chapter 7, we describe and discuss the measurement of the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$. In Chapter 8, we present a study of bias in techniques for measuring inclusive $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$ production in bottom jets. In Chapter 9, we summarize the results.

Chapter 2

The Theory of Hadron Production in e^+e^- Annihilation

In this chapter, we review the present understanding of the process $e^+e^- \rightarrow q\bar{q} \rightarrow hadrons$. We also discuss the implications of this understanding for the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$ and for the differences between bottom and non-bottom events. Finally, we review previous measurements of hadronization for events produced by different kinds of quarks.

2.1 Quarks and Their Properties

In order to understand the process $e^+e^- \rightarrow q\bar{q} \rightarrow hadrons$, we must first be familiar with some of the properties of quarks.

Matter is known to be composed of two classes of spin-1/2 particles: leptons and quarks. The distinction between these classes of particles is that while the quarks carry color charge and feel the strong force, the leptons do not. In the Standard Model, the particles are arranged into doublets [1] as shown in Figure 2.1. As far as we can tell, quarks and leptons are fundamental point particles with no sub-structure [2].

Some of the properties of the quarks are listed in Table 2.1. Quarks have never been observed in isolation and are apparently always confined within hadrons [4], so the masses of the quarks are not precisely known. The up, down, and strange quarks are referred to collectively as the light quarks, while the others are referred to collectively as the heavy quarks. The top quark is too heavy to be pair-produced at the center-of-mass energy of this experiment, 29 GeV, so when we refer to heavy quarks in this dissertation, we mean the charm and bottom quarks.

While it is conventional to refer to quarks only by their type, or *flavor*, each

Leptons Quarks

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \qquad \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

Figure 2.1: The Standard Model fermions (From [3]).

name	abbreviation	electric charge	mass
down	d	$-1/3$ e	9.9 ± 1.1 MeV/c^2 [5]
up	u	$+2/3$ e	5.6 ± 1.1 MeV/c^2 [5]
strange	s	$-1/3$ e	199 ± 33 MeV/c^2 [5]
charm	c	$+2/3$ e	1.35 ± 0.05 GeV/c^2 [5]
bottom	b	$-1/3$ e	$\sim$5 GeV/c^2 [6]
top	t	$+2/3$ e	$174 \pm 10^{+13}_{-12}$ GeV/c^2 [7]

Table 2.1: The properties of quarks.

quark listed in Table 2.1 actually represents three different quarks which are identical except that each has a different color (red, green, or blue). Likewise, antiquarks have one of 3 anticolors (anti-red, anti-green, or anti-blue), and gluons one of the 8 color-octet combinations of a color and an anticolor. Quantum Chromodynamics (QCD) is the theory of the color (or strong) force [8], in which gluons mediate the interaction between particles that possess a color charge. It is widely believed, but not proven, that the confinement of quarks and gluons within colorless hadrons is a property of QCD.

As far as we know, there are two ways of forming color-neutral hadrons from quarks. Mesons are made of a quark and an antiquark of the same color, such as red and anti-red. Examples of mesons are the π^+ (charged pion), with quark content $u\bar{d}$ and mass 139.6 MeV/c^2, and the K$^+$ (charged kaon), with quark content $u\bar{s}$ and mass 493.7 MeV/c^2. Baryons are a color-singlet combination of 3 quarks, each of which has a different color. The proton, with quark content uud and mass 938.3 MeV/c^2, is a baryon. Table 2.2 lists some of the most commonly produced hadrons containing heavy quarks.

2.2 Production of $q\bar{q}$ in e$^+$e$^-$ Annihilation

In the first step in the process e$^+$e$^-$ $\rightarrow q\bar{q} \rightarrow$ *hadrons*, an electron and a positron annihilate into a virtual photon, which couples into a final state consisting of a quark and its antiquark. Quarks are fundamental spin-1/2 fermions, therefore the first approximation to the cross-section for e$^+$e$^-$ $\rightarrow q\bar{q}$, for a quark q with electric charge $Q_q e$, has the usual cross-section for massless fermions in Quantum Electrodynamics (QED), the quantum theory of the electromagnetic force mediated by the photon [12]. This cross-section is

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4s}\left[1+\cos^2\theta\right]\ Q_q^2\ , \tag{2.1}$$

$$\sigma = \frac{4\pi\alpha^2}{3s}\ Q_q^2 \tag{2.2}$$

[13], where s is the square of the center-of-mass energy of the e$^+$e$^-$ system and $\alpha = e^2/\hbar c \simeq 1/137$ is the dimensionless coupling strength of QED.

In this approximation, we compute

$$\begin{aligned}\sigma(b\bar{b}) &= 3\,\frac{4\pi\alpha^2}{3s}\left(\frac{1}{3}\right)^2 = 0.0344 \text{ nb},\\ \sigma(q\bar{q}) &= 3\,\frac{4\pi\alpha^2}{3s}\left[3\left(\frac{1}{3}\right)^2 + 2\left(\frac{2}{3}\right)^2\right] = 0.3787 \text{ nb},\end{aligned} \tag{2.3}$$

name	quark content	spin	mass (GeV/c^2)
D^+	$c\bar{d}$	0	1.869
D^0	$c\bar{u}$	0	1.865
D_s^+	$c\bar{s}$	0	1.969
D^{*+}	$c\bar{d}$	1	2.010
D^{*0}	$c\bar{u}$	1	2.007
D_s^{*+}	$c\bar{s}$	1	2.110
Λ_c^+	cdu	1/2	2.285
Σ_c^{++}	cuu	1/2	2.453
Σ_c^+	cdu	1/2	2.453
Σ_c^0	cdd	1/2	2.453
Ξ_c^+	csu	1/2	2.466
Ξ_c^0	csd	1/2	2.473
B^0	$b\bar{d}$	0	5.279
B^-	$b\bar{u}$	0	5.279
B_s^0	$b\bar{s}$	0	5.375
B^{*0}	$b\bar{d}$	1	5.325
B^{*-}	$b\bar{u}$	1	5.325
B_s^{*0}	$b\bar{s}$	1	5.422
Λ_b^0	bdu	1/2	5.64

Table 2.2: Some hadrons containing heavy quarks [9, 10, 11].

and

$$\sigma(b\bar{b})/\sigma(q\bar{q}) = 1/11 \approx 0.0909 \ . \tag{2.4}$$

The overall factor of 3 in the cross-sections comes from the fact that each quark comes in 3 colors.

This is the lowest order approximation for the cross-sections. A discussion of higher-order effects in the calculation of these cross-sections is given in Section 2.7.

2.3 QCD in e^+e^- Annihilation

The quark and antiquark produced in e^+e^- annihilation, being colored objects, can radiate gluons, just as electric charges radiate photons. In QCD, quarks and gluons interact with a strength parameterized by a dimensionless coupling constant $\alpha_s(\sqrt{Q^2})$ that is a function of $Q^2 = -q^2$, where q_μ is the gluon 4-momentum. α_s is said to *run* with $\sqrt{Q^2}$. The *leading logarithm approximation* to this dependence is

$$\alpha_s\left(\sqrt{Q^2}\right) = \frac{12\pi}{(33 - 2N_f)\ln(Q^2/\Lambda^2)} \ , \tag{2.5}$$

where N_f is the number of quark flavors with mass less than $\sqrt{Q^2}$, and Λ is an experimentally measured constant[1] [14]. As $\sqrt{Q^2}$ decreases, $\alpha_s(\sqrt{Q^2})$ increases: $\alpha_s(91\text{ GeV}) = .115 \pm .008$ [15] and $\alpha_s(34\text{ GeV}) = .14 \pm .02$ [16]. When $\sqrt{Q^2}$ is in the neighborhood of 1 GeV/c, $\alpha_s(\sqrt{Q^2})$ becomes of order 1. As long as $\alpha_s(\sqrt{Q^2})$ is much less than 1, perturbation theory can be used to calculate cross-sections.

We now review the two approaches to perturbative QCD calculations: fixed order, and the leading logarithm approximation.

2.3.1 Fixed Order Computations in QCD

These computations are for final states composed of a small, definite number of partons (a parton is a quark, an antiquark, or a gluon). An example of a fixed order perturbative computation is the total cross-section for e^+e^- annihilation into a maximum of 5 partons at center-of-mass energy $\sqrt{s}$. This computation shows that

$$\sigma = \sigma(e^+e^- \to q\bar{q})\left[1 + \frac{\alpha_s(\sqrt{s})}{\pi} + 1.411\left(\frac{\alpha_s(\sqrt{s})}{\pi}\right)^2 - 12.8\left(\frac{\alpha_s(\sqrt{s})}{\pi}\right)^3\right] \tag{2.6}$$

[1] As $\sqrt{Q^2}$ decreases, so does N_f, but $\alpha_s(\sqrt{Q^2})$ is a continuous function of $\sqrt{Q^2}$, so each range of constant N_f actually has its own constant $\Lambda^{(N_f)}$, and all the $\Lambda^{(N_f)}$'s are related by Equation 2.5 and the continuity of $\alpha_s(\sqrt{Q^2})$.

for center-of-mass energy well above the $b\bar{b}$ threshold and well below the Z^0 mass [17, 18]. The Z^0 and the $W^{\pm}$ are the mediators of the weak force, which is described by the Weinberg-Salam model [19].

Another example of a fixed order perturbative computation is the differential cross-section for e^+e^- annihilation into a quark with a fraction $x_q = E_q/E_{\text{beam}}$ of the total energy, an antiquark with a fraction $x_{\bar{q}} = E_{\bar{q}}/E_{\text{beam}}$ of the total energy, and a gluon with a fraction $x_g = E_g/E_{\text{beam}}$ of the total energy [20]. The cross-section is

$$\frac{d\sigma}{dx_q dx_{\bar{q}}} = \sigma(e^+e^- \to q\bar{q})\frac{2\alpha_s(\sqrt{s})}{3\pi}\frac{x_q^2 + x_{\bar{q}}^2}{(1-x_q)(1-x_{\bar{q}})} \,. \tag{2.7}$$

This cross-section diverges for zero-energy gluons ($x_q = 1$ and $x_{\bar{q}} = 1$), for gluons colinear with the quark ($x_{\bar{q}} = 1$), and for gluons colinear with the antiquark ($x_q = 1$). These divergences (poles) cancel with the divergences in diagrams where a gluon is emitted and reabsorbed by the same quark or antiquark.

2.3.2 The Leading Logarithm Approximation of QCD

The other approach to perturbative QCD is *leading logarithm QCD*. This approach sums up, to all orders, the most divergent processes at each order in perturbation theory. The divergences are the colinear singularities of the type found in Equation 2.7.

Leading logarithm QCD allows us to model the perturbative evolution of an event as a series of the independent parton branchings $q \to qg$, $g \to q\bar{q}$, and $g \to gg$, with the probability of each branching given by one of the Altarelli-Parisi splitting functions [21]. The partons produced in each branching have lower virtuality than the original parton; the branching is stopped when the parton masses approach the energy scale Q_0 where perturbation theory fails. The entire branching process, illustrated in Figure 2.6a, is called a *parton shower*.

2.3.3 Where Perturbative QCD Fails

As $\sqrt{Q^2}$ decreases during the evolution of a hadronic event, $\alpha_s(\sqrt{Q^2})$ increases until it becomes of order 1, at about 1 GeV/c. At this energy scale, perturbation theory fails. In this energy region, where confinement and hadronization occur, we must resort to other means for calculating amplitudes and cross-sections for hadron production.

Lattice computations of QCD are first-principles computations of QCD, but require an enormous amount of computation, so only now are we beginning to obtain useful results on the simplest of lattice calculations [22, 23]. Detailed lattice QCD results on hadronization are still years away because of their complexity. As

of yet, there are no other methods for computing from first-principles QCD at low momentum transfer. Instead, we resort to phenomenological models of the hadronization process.

2.4 Models of Hadronization

There are a number of models of hadronization. Their Monte Carlo implementations all start with the production of a set of partons, either by fixed order QCD or by a parton shower. The hadronization model then transforms the parton configuration into a set of hadrons. Finally, in the Monte Carlo implementation, those hadrons with relatively short lifetimes are decayed, producing a set of particles that live long enough to travel an observable distance. This set of particles is then compared to experimental data.

In this analysis, hadronization models are used to compute acceptances, test the analysis method of Chapter 8 for bias (Sections 8.6 and 8.5), and train the neural networks we use to distinguish bottom events from non-bottom events (Section 7.1) and 8.3).

The hadronization models we describe in this section have Monte Carlo implementations that we can use for all of these purposes. These models can be grouped into three classes: independent fragmentation, string fragmentation, and cluster fragmentation. We then discuss the Peterson fragmentation function that can be used in those models with a fragmentation function: independent and string fragmentation.

2.4.1 Independent Fragmentation

Historically, one of the first fragmentation models was the independent fragmentation model of Feynman and Field [24]. In this model, the original quark and antiquark each transform into a jet of hadrons, independently of each other.

Figure 2.2 illustrates the creation of a jet by independent fragmentation from a quark q_0 created in the process $e^+e^- \rightarrow q_0\bar{q}_0$. First, a quark pair $q_1\bar{q}_1$ is created from the vacuum. $\bar{q}_1$ and q_0 combine to form a meson, leaving behind q_1, which has less energy than q_0 did. Then another pair $q_2\bar{q}_2$ is created, and q_1 and $\bar{q}_2$ bind together to form another meson, leaving behind q_2 with still less energy. This process repeats itself until the remaining quark has too little energy to form a meson. The same kind of iterative process produces a second jet from $\bar{q}_0$.

It is assumed that the quarks and antiquarks created from the vacuum each have a transverse momentum that is distributed as a Gaussian with an experimentally determined width σ_q. The total transverse momentum of each created pair is zero. Another experimentally measured parameter r determines the fraction of

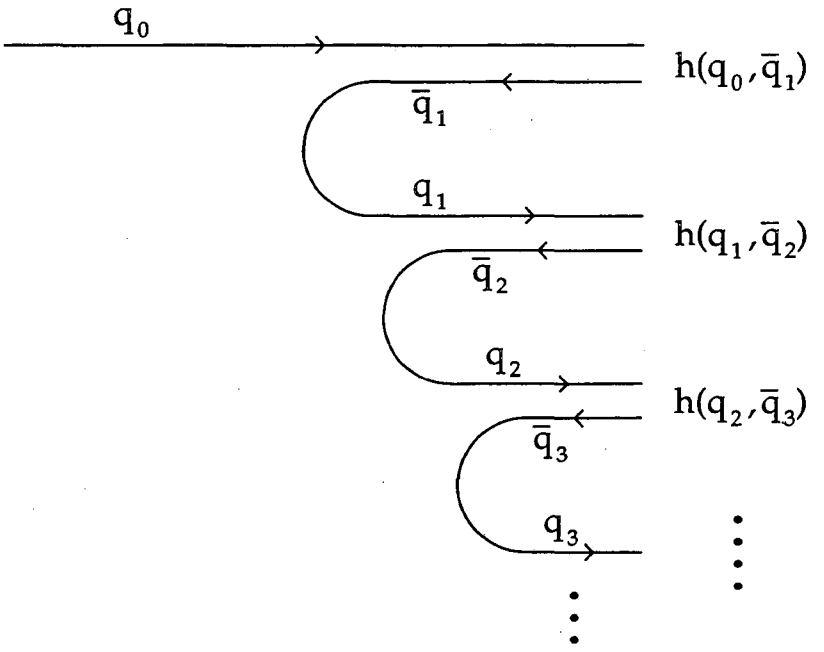


Figure 2.2: The independent fragmentation process into mesons for a jet initiated by the quark q_0. 'h(q_m,$\bar{q}_n$)' is a meson with quark content $q_m\bar{q}_n$ (Based on [25]).

mesons that are vector, the remainder being pseudoscalar. A third adjustable parameter, $P(s)/P(u)$, determines the probability that the created $q\bar{q}$ pairs are $s\bar{s}$, the remaining pairs being half $u\bar{u}$ and half $d\bar{d}$.

Another feature of this model is a *fragmentation function* $f(z)$. It is the probability density, at each step in the fragmentation chain, that a fraction z of the momentum of quark q_i goes into the meson formed by it and $\bar{q}_{i+1}$. The remaining momentum goes into q_{i+1}. Feynman and Field chose

$$f(z) = 1 - a - 3a(1-z)^2 \ , \tag{2.8}$$

where a is determined from experiment.

Some of the hadrons created in this cascade (called primary hadrons) are unstable and are decayed by the Monte Carlo implementation of the Feynman-Field model. The order in which these primary hadrons are created is called the *rank*; the first primary hadron has rank 1, the second one has rank 2, etc.

The original Feynman-Field model did not include baryon production and gluon jets. Meyer [26] proposed an extension of the model in which occasionally two quark-antiquark pairs, rather than one, are created from the vacuum with probability $P(qq)/P(q)$. The qq then combines with the adjacent quark, and the $\bar{q}\bar{q}$ combines with the adjacent antiquark, forming a baryon-antibaryon pair.

Hoyer [27] and Ali [28] introduced gluon jets, which split into $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ pairs with equal probability. The quark and antiquark then fragment independently. Different variations on independent fragmentation divide the gluon momentum between the quark and antiquark differently. Massive quarks must be handled differently from light quarks; it was in this context that the Peterson function was invented (Section 2.4.4).

Even though independent fragmentation is basically a parameterization, rather than an attempt to model the fundamental dynamics of fragmentation, it is effective in describing hadron production. This model simulates the jets in events independently of each other, so we use it in the studies in Section 8.6.

It is not the model of choice, however. It fails to reproduce the string effect [29, 30, 31]. Also, it has a number of serious theoretical problems. There is no natural way of handling the last (anti)quark in each jet; some variations on independent fragmentation simply throw them away. Neither energy nor momentum is conserved, unless the jets are rescaled in E and/or p in an ad hoc manner [32]. Finally, since the properties of the cascade depend upon the initial (anti)quark momentum in the laboratory frame of reference, independent fragmentation is not Lorentz covariant.

2.4.2 String Fragmentation

Another, more physical model of fragmentation is string fragmentation, proposed by Artru and Mennessier [33] and Andersson [34, 35, 36]. In this model, the color field between a quark and an antiquark is a massless color flux tube, or string, that is uniform along its length. The popular implementation of this model is the Lund model [36]. Jetset is the name of the software package in which the Lund model is implemented.

Several facts suggest that the color flux forms a uniform string. An electric dipole has the familiar configuration of Figure 2.3a that spreads out to infinity: as the charges separate, the field lines spread out, yielding a force that is the inverse square of the separation r between the charges. Photons do not carry charge, but gluons carry color since the strong force is non-Abelian, so it is plausible that color flux lines attract one another, constricting the dipole pattern and producing a force that falls off less rapidly than $1/r^2$. That the flux lines form a flux tube, as shown in Figure 2.3b, is suggested by the linearity of Regge trajectories [37], Lattice QCD [38], and the long-distance behavior of QCD potential models [39].

The flux tube is uniform along its length; therefore it has constant energy per length κ, which is experimentally measured to be about 1 GeV/fm. The force between a quark and antiquark joined by such a string is independent of their separation.

The string model of a meson of mass m at rest, composed of a massless quark and a massless antiquark, is illustrated in Figure 2.4. At time $t = 0$, the quark and antiquark are moving apart at the speed of light. As the quark and antiquark separate, their energy goes into the string until the string energy is equal to the mass of the meson and the quarks have no energy, at which point both quark and antiquark turn around and move toward each other and eventually past each other at the speed of light. The cycle repeats itself. Let us label the space-time points at which the quark and antiquark turn around be (x_1, t_1) and (x_2, t_2). Then it is true that $t_2 - t_1 = 0$ and $x_2 - x_1 = m/\kappa$. Thus we obtain the expression

$$(x_2 - x_1)^2 - c^2(t_2 - t_1)^2 = m^2/\kappa^2 \ . \tag{2.9}$$

This equation is Lorentz invariant. If the meson has transverse momentum, then m in the above expression is replaced by the *transverse mass* $m_\perp = \sqrt{p_\perp^2 + m^2}$. This is the so-called yo-yo model of a meson.

Hadronization of a quark and an antiquark created in e^+e^- annihilation at high energy, as described by the Lund model (Figure 2.5), shares many features with the yo-yo meson model. The quark and antiquark are produced moving apart with a string connecting them. As the quark and antiquark separate, energy goes into the string and eventually it is energetically advantageous to break the string with the creation of quark-antiquark pairs from the vacuum; these pairs terminate the

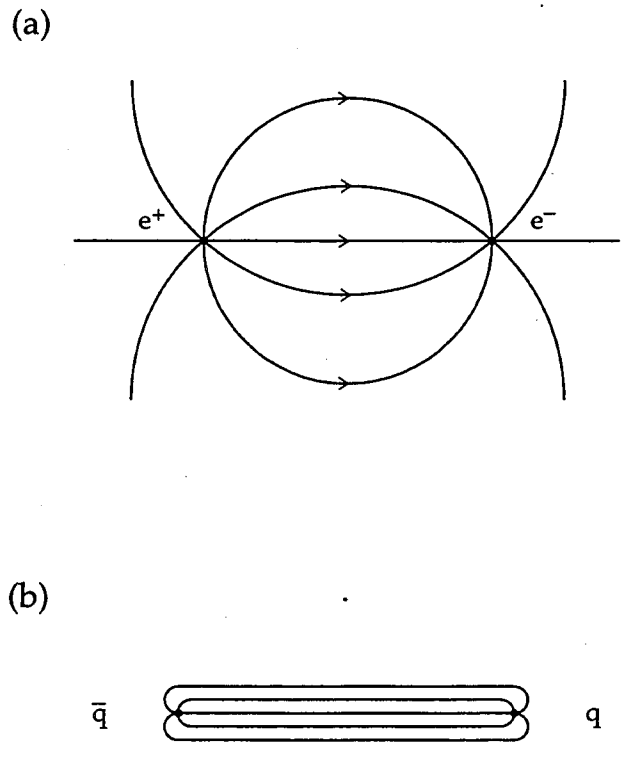


Figure 2.3: Electric flux lines for a static electric dipole (a). Color flux lines for a $q\bar{q}$ pair (b) (Based on [25]).

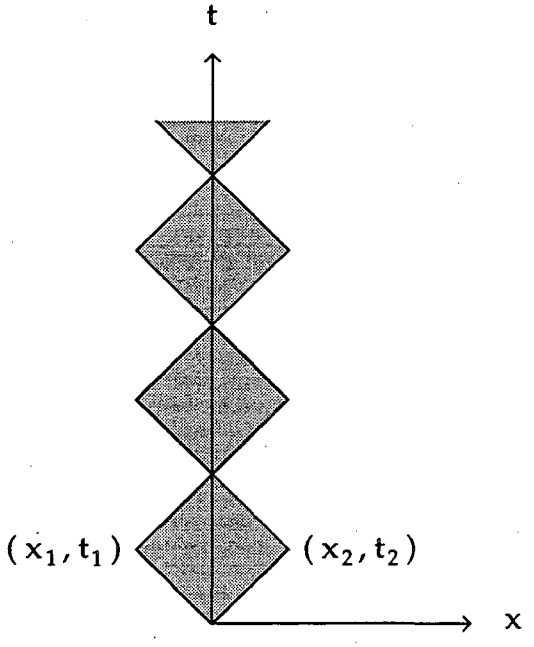


Figure 2.4: The yo-yo model of a meson in string fragmentation (Based on [25]).

flux lines and thus break the string. The string will break a number of times until there is not enough energy to create new quark antiquark pairs, at which point the fragmentation of the system is complete, and each quark-antiquark pair along with the string that connects them forms a meson (the little yo-yos in Figure 2.5). Note that Equation 2.9 applies to the creation points of the $q\bar{q}$ pairs that break the string, since the mesons that are created in the process must be on-shell.

The creation of a quark-antiquark pair at a point violates energy conservation if the quarks have mass or transverse momentum. Once the quark and antiquark have separated by a distance $d = m_\perp/\kappa$, energy conservation is restored. Thus, the breaking of a string by a $q\bar{q}$ pair separated by a distance d is a tunneling event, which has a probability of occurring given in quantum mechanics by

$$P \propto \exp\left(\frac{-\pi m_\perp^2}{\kappa}\right) . \tag{2.10}$$

As a consequence of this equation, the probability differs for creating different kinds of quark pairs in breaking the string. If we use $\kappa = 1$ GeV/fm $= 0.2$ GeV^2 and the quark masses $m_u = m_d = 0$, $m_s = 250$ MeV/c^2 and $m_c = 1.5$ GeV/c^2, then the relative probabilities for the production of up, down, strange, and charmed quarks are approximately $1:1:0.37:10^{-16}$. Thus, strange quark production is suppressed, and heavier quarks basically are not produced at all in breaking the string. In the Lund model, the strange quark production probability is left as a free parameter, as in independent fragmentation. Similarly, the probability to produce vector versus pseudoscalar mesons, as well as the probability to produce diquarks for the production of baryons, are left as free parameters to be determined experimentally. The transverse momenta of the quark and antiquark are also equal and opposite, and their distribution follows a Gaussian distribution with width σ_q.

The longitudinal momenta, p_l, of these hadrons are determined by a fragmentation function, as in independent fragmentation, but the string fragmentation function is a function of the light-cone variable $E + p_l$ instead of p_l. The first hadron created at the end of the string takes up a fraction ζ_1 of the total $E + p_l$ of the entire string:

$$(E + p_l)_1 = \zeta_1 (E + p_l)_{total} . \tag{2.11}$$

$E + p_l$ of the next hadron created is another fraction ζ_2 of the remaining available $E + p_l$ of the unfragmented string system. This procedure is repeated until a certain minimum $E + p_l$ is reached. At this point the remaining string decays into two mesons according to two-body phase space. An important feature of the Lund model is the requirement that, on average, the fragmentation starting from one end of the string is the same as the fragmentation starting from the other end of the string. This determines the form of the *Lund Symmetric fragmentation function* [40]:

$$f(\zeta) \propto \frac{(1-\zeta)^a}{\zeta} \exp\left(\frac{-bm_\perp^2}{\zeta}\right) . \tag{2.12}$$

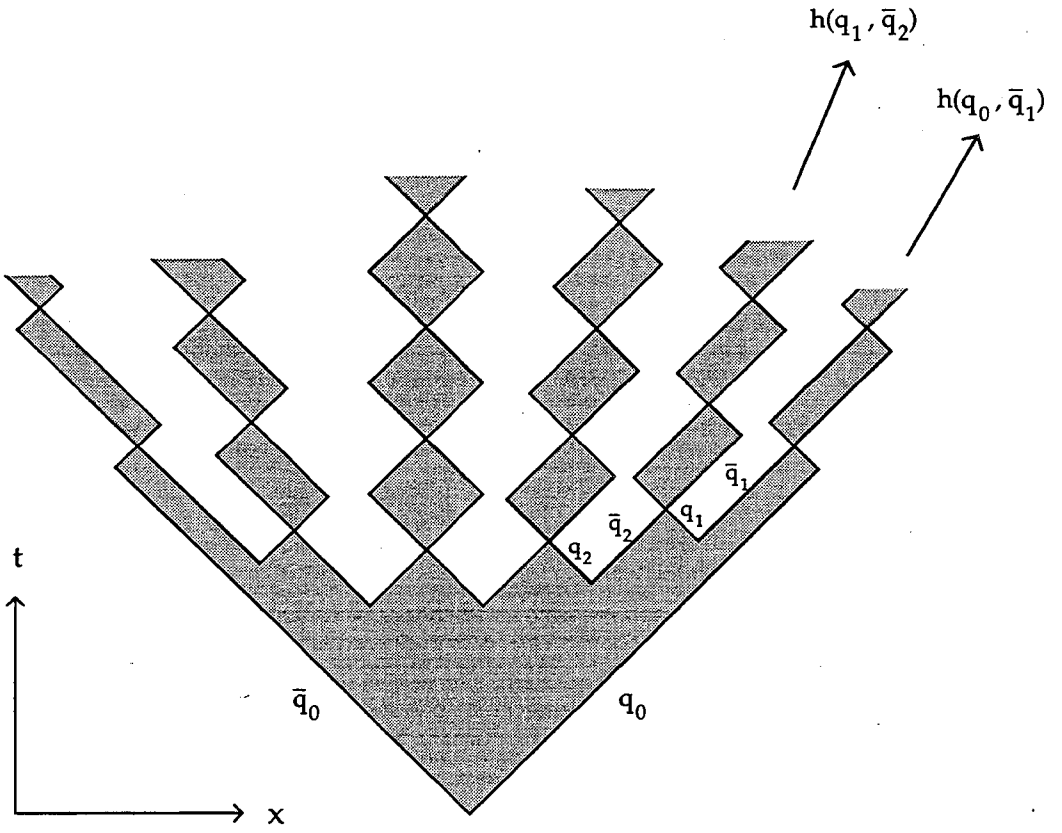


Figure 2.5: String fragmentation into mesons at high energy. The shaded areas are the regions where the flux tube exists, and the solid lines are the quark and antiquark trajectories (Based on [25]).

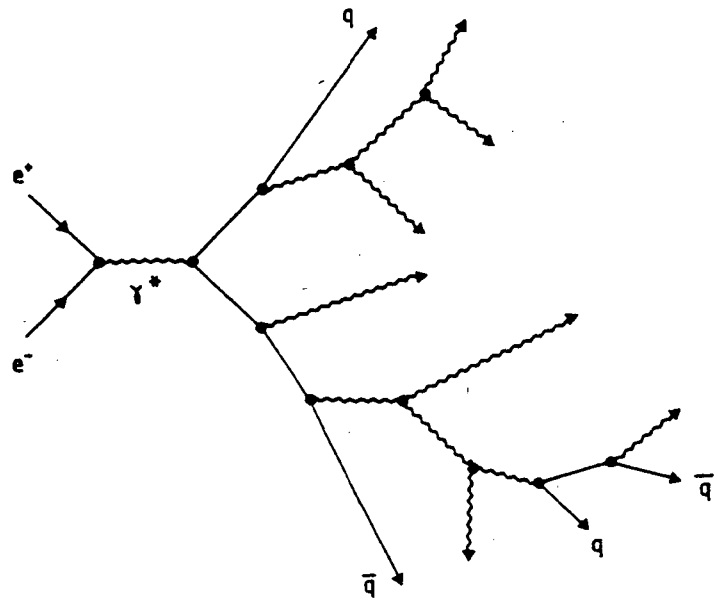
Another important feature of the Lund model is that gluons are accommodated as kinks in the string. In older versions of the Jetset package, these gluons were generated according to second order QCD. In more recent versions of Jetset (6.3 or higher-numbered versions), it is also possible to generate gluons using a leading logarithm parton shower. The Lund model does not suffer from problems with Lorentz covariance and E-p conservation, but it has many parameters. The Lund model reproduces experimental data well [41], and we use it to train the neural networks we use to distinguish bottom events from non-bottom events (Sections 7.1 and 8.3).

2.4.3 Cluster Fragmentation

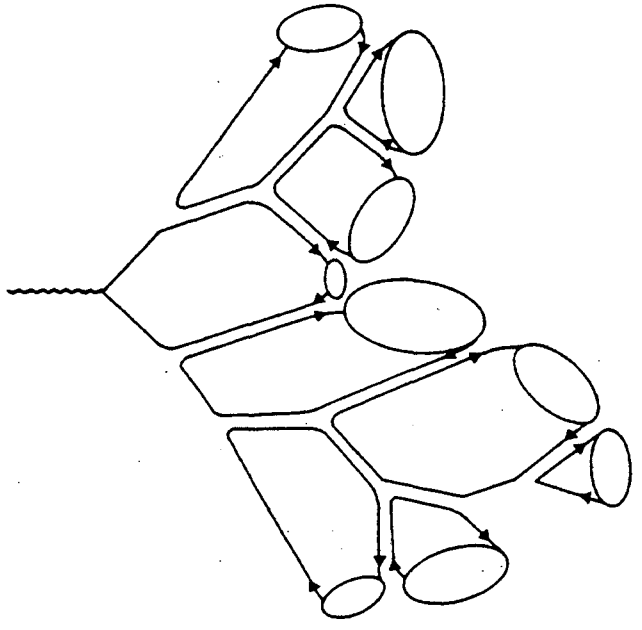
The third class of fragmentation models used in high energy physics are the cluster fragmentation models. In this approach, a leading logarithm parton shower is generated and the evolution of the shower is terminated when the parton virtuality Q falls below a cutoff Q_0. At this point all gluons are split into quark-antiquark pairs, and adjacent quarks and antiquarks are formed into colorless clusters [42] (Figure 2.6). These clusters then decay into pairs of hadrons according to two-body phase space [43].

The only other really fundamental parameter in cluster models besides Q_0 is the QCD scale parameter Λ, which determines the strength of the strong force. Parameters for determining strange quark production, baryon production, and a transverse momentum distribution are all unnecessary in cluster models, since all these characteristics are taken care of by the shower cutoff scale. The strange quark and baryon production fractions are limited by available phase space, and transverse momentum is governed by the masses of the created clusters. The most popular cluster fragmentation model, the Webber model [44], has a few additional parameters. One of these, M_f, is used to fission clusters that are very massive: these clusters are fissioned using a string mechanism. The Webber model also has parameters for quark masses. Up and down quark masses are fixed to be $Q_0/2$, and the other quarks are given their usual masses.

Note that cluster fragmentation has no fragmentation function; the shapes of the hadron momentum spectra are determined by the shower and cluster processes. In spite of the far fewer parameters in the Webber model, it reproduces data well [41]. We do not use it in this analysis, since the charm and bottom hadron momentum spectra can not be independently tuned to match experimental data.



(a)



(b)

Figure 2.6: Cluster fragmentation. (a) A leading logarithm parton shower, and (b) the color flow of this shower after gluons have split into $q\bar{q}$ pairs and color has been confined in clusters, which are represented by the ellipses (From [44]).

2.4.4 The Peterson Fragmentation Function

The Field-Feynman function predicts heavy quark fragmentation that gives charm and bottom hadrons too little momentum [45], while the Lund Symmetric fragmentation function can not get charm, bottom, and light-quark fragmentation all correct at the same time [46]. Left as is, this is a fatal flaw for any simulation of bottom quark event behavior.

The Peterson function [47], a simple fragmentation function, allows tuning of the Monte Carlo charm and bottom momentum spectra to match the experimentally measured spectra. It describes accurately the shape of heavy hadron momentum spectra [48], has only one free parameter for each heavy quark (that must be experimentally measured), and has a simple derivation based upon the uncertainty principle and longitudinal phase space. The Peterson function is the standard in bottom event analyses because of these facts.

We now show a derivation of the form of the Peterson function. The transition amplitude from a heavy quark Q with momentum p to a hadron H (quark content $Q\bar{q}$) with momentum zp plus a light quark q with momentum $(1-z)p$ is given in first order perturbation theory by

$$\frac{\langle Hq|\mathcal{H}'|Q\rangle}{\Delta E}\,, \tag{2.13}$$

where $\Delta E = E_Q - E_H - E_q$ and $\mathcal{H}'$ is the perturbing Hamiltonian. The fragmentation function is then given by the square of the transition amplitude

$$f(z) \propto \frac{1}{z(\Delta E)^2}\,, \tag{2.14}$$

where the factor of $1/z$ comes from longitudinal phase space. If the Q and q are moving rapidly, then we can approximate

$$E_Q = \sqrt{M_Q^2 + p^2} = p\sqrt{1 + (M_Q/p)^2} \simeq p + \frac{M_Q^2}{2p} \tag{2.15}$$

$$E_H = \sqrt{M_H^2 + z^2p^2} \simeq zp + \frac{M_H^2}{2zp} \tag{2.16}$$

$$E_q = \sqrt{M_q^2 + (1-z)^2p^2} \simeq (1-z)p + \frac{M_q^2}{2(1-z)p}\,, \tag{2.17}$$

implying that

$$\begin{aligned} f(z) &\propto \frac{1}{z\left[1 - \left(\frac{M_H}{M_Q}\right)^2/z - \epsilon_Q/(1-z)\right]^2} \\ &\approx \frac{1}{z[1 - 1/z - \epsilon_Q/(1-z)]^2}\,, \end{aligned} \tag{2.18}$$

with $\epsilon_Q = (M_q/M_Q)^2$ the so-called Peterson parameter.

In Figure 2.7, we show the Peterson function as tuned in this analysis for charm ($\epsilon_c = 0.072$) and bottom ($\epsilon_b = 0.039$) hadrons in Jetset 7.2. The tuning process is described in Section 5.3.2.

The interpretation of z varies. In this dissertation, we adopt the most common interpretation that z is $(E + p_{\|})_{\text{hadron}}/(E + p)_{\text{quark}}$, where the parallel direction is with respect to the quark direction. z involves unobservable kinematic quantities of quarks, so z must be inferred from a fragmentation model, and ϵ must be tuned individually for each Monte Carlo fragmentation model implementation. For Jetset 5.2, it was found that $\epsilon_c = 0.06$ and $\epsilon_b = 0.006$ [49]. The values of the Peterson parameters for Jetset 5.2 and 7.2 are very different because Jetset 5.2 uses a 2nd-order computation of QCD, while Jetset 7.2 uses a leading logarithm computation of QCD, and these two different computations produce different relationships between $(E + p_{\|})_{\text{hadron}}$ and $(E + p)_{\text{quark}}$.

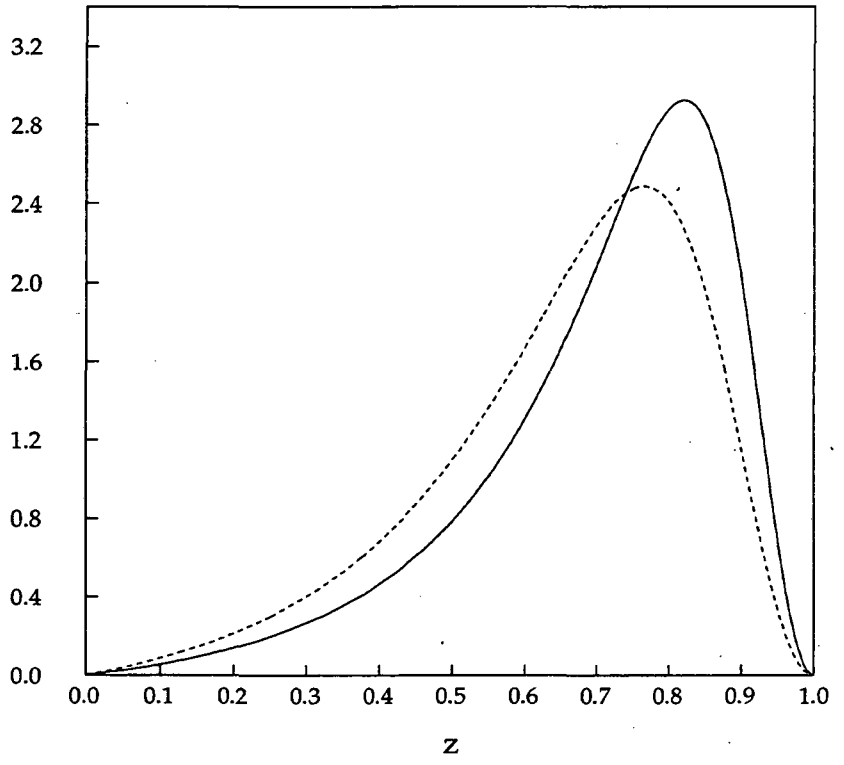


Figure 2.7: The Peterson fragmentation function as tuned in this analysis for charm hadrons (dashed) and bottom hadrons (solid).

2.5 The Properties of Heavy Quark Events

In Table 2.3, we show some properties of the charged particles in jets that originated from different types of $q\bar{q}$ pairs created at $\sqrt{s} = 29$ GeV. The way in which these properties were measured is reviewed in Section 2.6. Within the errors on the measurements, light-quark and charm-quark jets have the same properties. Bottom quark jets, in contrast, have significantly higher charged multiplicity and lower average particle momenta, and are rounder, than other types of jets. However, the average tranverse momentum (p_t or $p_\perp$) of charged tracks, with respect to the event axis, is the same in bottom and non-bottom events; the reason for this is discussed below.

The large mass of the bottom quark, about 5 GeV/c^2, is responsible for these differences. This large mass causes bottom hadrons to take up a large fraction of the energy of the primary $b\bar{b}$ in bottom events at $\sqrt{s} = 29$ GeV [50, 51]. The energy fraction has been measured to be, on average, about 72% (Section 5.3.2). Of the remaining 28%, according to Monte Carlo, an average of 4% is lost to initial state radiation and an average of 24% goes into creating other hadrons. Eliminating ISR but keeping s the same, we assume that a fraction $72/(24+72)$ of the ISR photon's energy goes into the bottom hadrons and the rest goes into creating other hadrons. On average, 6.0 other charged hadrons are created per event[2]. The measured average charged multiplicity of a 50–50 mix of B^- and B^0 meson decays is 5.4 [54, 55], and since bottom hadrons do not vary widely in mass (see Table 2.2), we assume that the average charged multiplicity of the mix of primary bottom hadrons created at $\sqrt{s} = 29$ GeV is the same. Therefore, we predict that the average charged multiplicity of bottom jets is roughly $6.0/2 + 5.4 = 8.4$, which is about 2 σ above the measured average bottom jet multiplicity of 7.8 and much larger than the average event charged multiplicity of 6.2 (see Table 2.3[3]).

This larger average multiplicity causes the event energy to be shared among more particles. Therefore, the average momentum, parallel to the jet direction, of particles in bottom jets is smaller than in non-bottom jets. This effect is enhanced by the fact that bottom hadrons tend to share their decay energy equally among the larger number of its decay products, whereas in other events, particles are ordered in rank and the primary hadron containing the quark that initiated the jet will generally be the fastest primary hadron.

Bottom hadron masses are much greater than the masses of the charm hadrons they can decay into. Therefore, the decay of bottom hadrons releases a lot of

[2]We use the parameterization that the dependence of the average charged multiplicity $\langle n_{ch} \rangle$ on s is $3.24 - 0.34\ln(s) + 0.26\ln^2(s)$ [52], and we make the common assumption that this parameterization holds for the non-bottom-hadron portion of the event. This assumption appears to hold [53].

[3]The multiplicities in this table are for events with no ISR, which is why we have eliminated ISR in this estimate of the multiplicity.

energy, giving its decay products, as a group, a larger momentum transverse to the jet direction than is available in other events. As a result, bottom quark jets are fatter and rounder[4].

The average p_t of particles in bottom jets is, within statistics, the same as for particles in non-bottom jets, since the greater net transverse momentum is distributed among more particles, and the two effects apparently cancel each other out.

This cancellation does not happen for the leptons (electrons and muons) from the semi-leptonic decays of bottom hadrons, where the term semi-leptonic means the process $b \rightarrow c + W^-$ and $W^- \rightarrow \ell^- + \bar{\nu}_\ell$, or the charge conjugate process, with $\ell =$ e or μ. Leptons from the semi-leptonic decays of bottom hadrons often have a larger p_t than those from other sources, since they carry, on average, half the p_t of the W, which in turn carries half the p_t of the bottom decay products. Thus, the lepton p_t spectrum scales with the decaying hadron's mass. This large p_t is responsible for the usefulness of identifying, or *tagging*, bottom-quark events with hi-p_t leptons.

property	average jet	uds jet	c jet	b jet
multiplicity	6.23±.09	5.89±.24	6.60±.25	7.84±.29
$\langle p \rangle$, GeV/c	1.2935±.0015	1.52±.04	1.38±.06	1.06±.04
$\langle p_\perp \rangle$, GeV/c	.380±.005	.40±.01	.39±.01	—
$\langle p_\perp^2 \rangle$, (GeV/c)^2	.274±.001	—	—	.31±.03
$\langle S \rangle$	.1399±.0009	.087±.007	.094±.010	.26±.02
$\langle 1 - T \rangle$	.1059±.0003	.092±.004	.082±.005	.149±.009

Table 2.3: Properties of jets of different flavors [53, 56, 57, 58, 59].

2.6 Previously Used Heavy Quark Event Tags

The measurements that produced the entries in Table 2.3 are listed in Table 2.4. All use tags in one hemisphere to find the type of $q\bar{q}$ that originated the event, and the tracks in the other hemisphere are used for the measurement. For various reasons, these tags yield high-purity samples, but with low efficiency.

The high-p_t lepton tag [57, 58, 59] yields high-purity bottom event samples because other sources of leptons (semi-leptonic decays of charm hadrons in charm

[4] 3-jet events with an energetic gluon also have a lot of p_t, but these events are distinctly planar, since momentum conservation requires that the $q\bar{q}g$ initial state lies in a plane.

events, electrons from photon conversions, and muons from pion and kaon decays) generally do not have large p_t and large momentum. The physics of this tag was discussed in the last section. The efficiency of this tag is limited by several factors. The branching ratios for $b \rightarrow e + \mathrm{X}$ and for $b \rightarrow \mu + \mathrm{X}$ are only 10.5% [60]. A cut of 1 GeV/c on p_t is needed to beat down the background from semi-leptonic decays of primary charm hadrons, primary meaning not from decays of bottom hadrons. Finally, experiments can not use low-momentum electrons and muons, because of acceptance and low-momentum backgrounds from photon conversions and from pion and kaon decays [61].

The low-p_t lepton tag [58, 59] produces fairly pure samples of charm events because prompt leptons (i.e. electrons not from photon conversions and muons not from pion or kaon decays) are essentially always from the semi-leptonic decays of charm and bottom hadrons, because leptons from bottom hadrons tend to have larger p_t, and because there are 4 times as many charm events as there are bottom events. Its efficiency is limited by the roughly 13% branching ratios for $c \rightarrow e + \mathrm{X}$ and for $c \rightarrow \mu + \mathrm{X}$ [62], and by the typical inability of experiments to use low-momentum electrons and muons.

The $D^{*\pm}$ meson tag [53, 56] yields a highly pure sample of charm events because charm quarks hadronize into $D^{*\pm}$ 3/8ths of the time, and because few candidate $D^{*\pm}$ mesons are not $D^{*\pm}$. $D^{*\pm}$ mesons also come from bottom hadron decays, but these mesons have low momenta, whereas $D^{*\pm}$ mesons from hadronization of charm have higher momenta. The chosen cut on $x_E = E(D^{*\pm})/E_{\mathrm{beam}}$ for the TASSO measurement is 0.5 [53], and 0.4 and 0.5 for different parts of the HRS measurement [56]. The low efficiency of this tag is due to the fact that only a small fraction of the $D^{*\pm}$ mesons can be reconstructed.

The high-x_E charged kaon and pion tag [56], where x_E is $E_{\mathrm{particle}}/E_{\mathrm{beam}}$, produces high purity light-quark event samples because these events have a *leading* particle (i.e. containing the original quark or antiquark) that can be stable. The leading primary hadron is usually the fastest particle in its jet. When a particle decays, its x_E is shared among its decay products: decays feed down in x_E. Charm and bottom hadrons always decay, and with relatively high multiplicity, so charm and bottom events have very few particles with large x_E. As discussed in the last section, this difference between light- and heavy-quark jets is enhanced by the fact that hadrons containing a heavy quark tend to share their decay energy equally among their decay products, whereas in other events, particles are ordered in rank and the primary hadron containing the quark that initiated the jet will generally be the fastest primary hadron. The low efficiency of this tag is caused by the high cut on x_E (0.7 in the HRS analysis [56]) needed to eliminate tails from charm and bottom events and by the low probability for a hadron to have so much momentum.

All of these tags have low efficiency, causing the statistical significance of all

experiment	tag method	quark tagged	purity	efficiency
TASSO [53]	$D^{*\pm}$	c	80%	0.80%
HRS	$D^{*\pm}$	c	85%	0.60%
[56]	hi-x_E hadron	u,d,s	89%	1.14%
DELCO [57]	hi-p_t lepton	b	83%	1.86%
Mark II	hi-p_t lepton	b	64%	5.8%
[58]	lo-p_t lepton	c	35%	2.5%
TPC	hi-p_t lepton	b	65%	6.4%
[59]	lo-p_t lepton	c	52%	4.1%

Table 2.4: Previous hadronization measurements for different quark flavors.

these measurements to suffer. In an attempt to obtain greater statistical significance, this analysis uses a number of *hadronic variables* computed from the 3-momenta of all of the observed charged hadrons in the event or jet to statistically separate bottom (b) and non-bottom (non-b) events or jets. Each of these hadronic variables is well-defined and carries information for all events/jets, so this method is intrinsically high-efficiency. The trade-off is that the b and non-b distributions overlap in the entire range of all hadronic variables, and that the distributions are fairly similar. This is why more than one variable is used: to provide more information for distinguishing the two types of events/jets.

In this analysis, a neural network transforms all these hadronic variables into one variable that contains essentially all of the information in the hadronic variables that distinguishes b and non-b events/jets from each other [63]. This data compression makes the analysis considerably easier, since it is difficult to fit in a large number of variables. Neural networks are described in Chapter 6.

2.7 Corrections to $\sigma(b\bar{b})/\sigma(q\bar{q})$

The calculation of the theoretical value of the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$, to compare to the measurement of this ratio presented in this dissertation, requires making a number of corrections to the lowest order approximation of this ratio that was made in Section 2.2. There is often *initial state radiation* (ISR), a photon radiated by either the electron or the positron before they annihilate. The masses of the initially created quarks were ignored. Gluon emission by the quarks also alters the cross-sections. Finally, the annihilation photon contains a small admixture of Z^0.

It turns out that the heaviness of the bottom quark relative to the other produced quarks affects the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$. The differential cross-section for massive fermions is

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4s}\,\beta\,[1+\cos^2\theta+(1-\beta^2)\sin^2\theta]\,Q_q^2\,, \tag{2.19}$$

where $\beta = v/c$, and v is the quark's (and the antiquark's) speed in the center-of-mass (lab) frame [64]. The corresponding total cross-section is

$$\sigma = \frac{4\pi\alpha^2}{3s}\,\beta\,\left(\frac{3-\beta^2}{2}\right)\,Q_q^2\,, \tag{2.20}$$

or $\beta(3-\beta^2)/2 = .995$ times the massless QED cross-section in Equation 2.2, for b and $\bar{b}$ each with 14.5 GeV of energy and assuming that the bottom quark mass is 5 GeV/c^2. Therefore, the direct effect of the bottom quark mass is negligible.

In combination with initial state radiation, though, the bottom quark mass has a significant effect. Let $\sqrt{s}$ be the center-of-mass energy of the electron and positron before the emission of the ISR photon (the conventional definition), and let $\sqrt{s_{e^+e^-}}$ be the center-of-mass energy of the electron and positron after the emission of the ISR photon. The cross-section for e^+e^- annihilation after ISR photon emission is a variation on Equation 2.20:

$$\sigma = \frac{4\pi\alpha^2}{3s_{e^+e^-}}\,\beta_{e^+e^-}\,\left(\frac{3-\beta^2_{e^+e^-}}{2}\right)\,Q_q^2\,, \tag{2.21}$$

where $\beta_{e^+e^-}$ is β of the quarks in the boosted e^+e^- center-of-mass frame of reference after the emission of the ISR photon. For events with energetic ISR, $\beta_{e^+e^-}$ for bottom quarks can be significantly less than 1. If ISR is energetic enough, $\sqrt{s_{e^+e^-}}$ is below the threshold to produce a $b\bar{b}$ pair, and the virtual annihilation photon instead decays into pairs of the lighter quarks.

There is no analytic expression for the annihilation cross-section including the effects of initial state radiation. Instead, we use the Monte Carlo package described in Section 5.3.1, which contains a standard simulation of initial state radiation [65, 66] to calculate the total cross-section. We get

$$\begin{aligned} \sigma(b\bar{b}) &= 0.0413 \text{ nb} \\ \sigma(q\bar{q}) &= 0.5014 \text{ nb, and} \\ \sigma(b\bar{b})/\sigma(q\bar{q}) &= 0.0824\,, \end{aligned} \tag{2.22}$$

a significant change in the ratio from the value of .0909 found with the lowest order approximation in Equation 2.4.

Including the effects of QCD in our Monte Carlo, in addition to ISR, gives us

$$\begin{aligned} \sigma(b\bar{b}) &= 0.0437 \text{ nb} \\ \sigma(q\bar{q}) &= 0.5310 \text{ nb, and} \\ \sigma(b\bar{b})/\sigma(q\bar{q}) &= 0.0823 \,. \end{aligned} \tag{2.23}$$

Even though the cross-sections change significantly, QCD affects bottom and non-bottom events the same way, so the ratio of the cross-sections does not change.

The effect of the small admixture of Z^0 in the annihilation photon is quite minor. Added to the previous effects, we get

$$\begin{aligned} \sigma(b\bar{b}) &= 0.0438 \text{ nb} \\ \sigma(q\bar{q}) &= 0.5317 \text{ nb, and} \\ \sigma(b\bar{b})/\sigma(q\bar{q}) &= 0.0824 \,. \end{aligned} \tag{2.24}$$

This cross-section ratio is 90.6% of the ratio in Equation 2.4 that we started with.

Chapter 3

The TPC/Two-Gamma Experiment

In this chapter, we describe the experimental apparatus that collected the data we use to make the measurement presented in this dissertation.

This experimental apparatus, the TPC/2γ detector, was located in Interaction Region 2 of the PEP e^+e^- storage ring. PEP collided counter-rotating triads of bunches of electrons and positrons in six interaction regions (Figure 3.1). The particles in each bunch had an energy of 14.5 GeV, so the center-of-mass energy of each collision was 29 GeV.

The TPC/2γ detector was placed on the PEP ring in January, 1982, and it collected three data sets before it was shut down in September 1990. The first data set of 77 pb^{-1} was collected in 1982 and 1983; it is called the Low-Field data set because the detector had a conventional magnet that supplied a field of 4.0 kG. The second data set of 66 pb^{-1}, collected between 1984 and 1986, is called the High-Field data set because the conventional magnet had been replaced by a superconducting magnet that produced a field of 13.25 kG. The last data set of 32 pb^{-1}, collected in 1988 and 1990, is called the Vertex Chamber data set because of the Vertex Chamber that was used during this period. The Vertex Chamber replaced the Inner Drift Chamber that was previously part of the detector.

We use only the High-Field data set for the analysis presented in this dissertation, so we only describe the configuration of the detector during this period.

3.1 The TPC/Two-Gamma Detector

The TPC/2γ detector [68], shown in Figures 3.2, 3.3, and 3.4, is a 4π detector designed to gather information about charged and neutral particles in a large as possible solid angle. Arranged concentrically around the beam pipe, going from

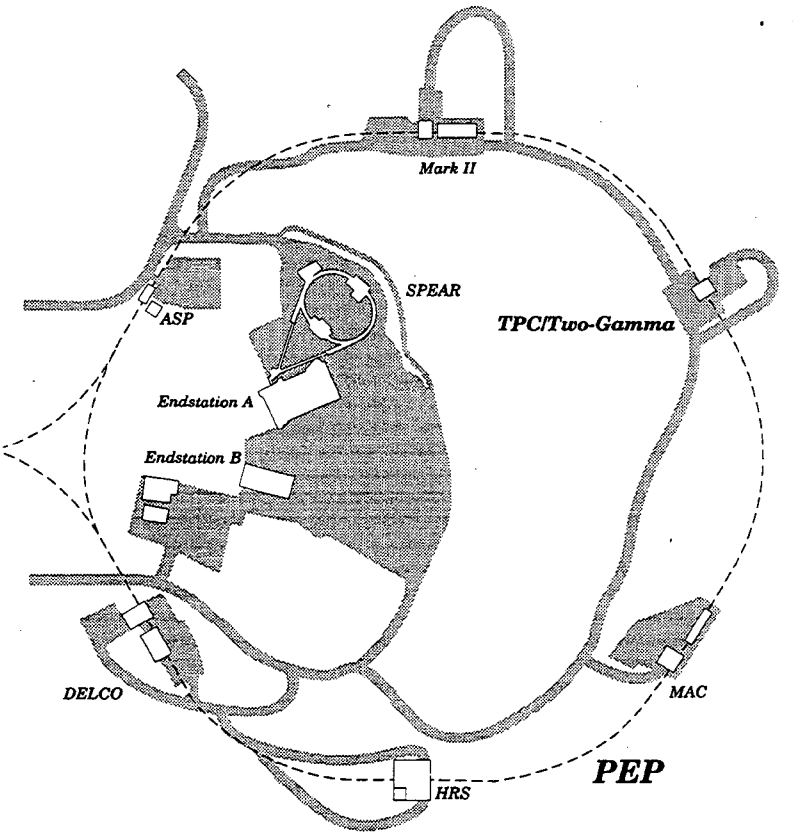


Figure 3.1: A schematic representation of the PEP storage ring. Electrons circulate in the clockwise direction, positrons counter-clockwise (From [67]).

the inside out, are the Inner Drift Chamber (IDC) [69, 70], the Time Projection Chamber (TPC) [71, 72], the Outer Drift Chamber (ODC) [70], a superconducting magnet coil, the hexagonal electromagnetic calorimeter (HEX) [73, 74], and the barrel muon chambers [75]. On the ends of these detectors, moving outwards from the TPC, are located the Pole Tip Calorimeter (PTC) [76], the Muon Doors [75], and the PEP-9 (forward) detectors [77].

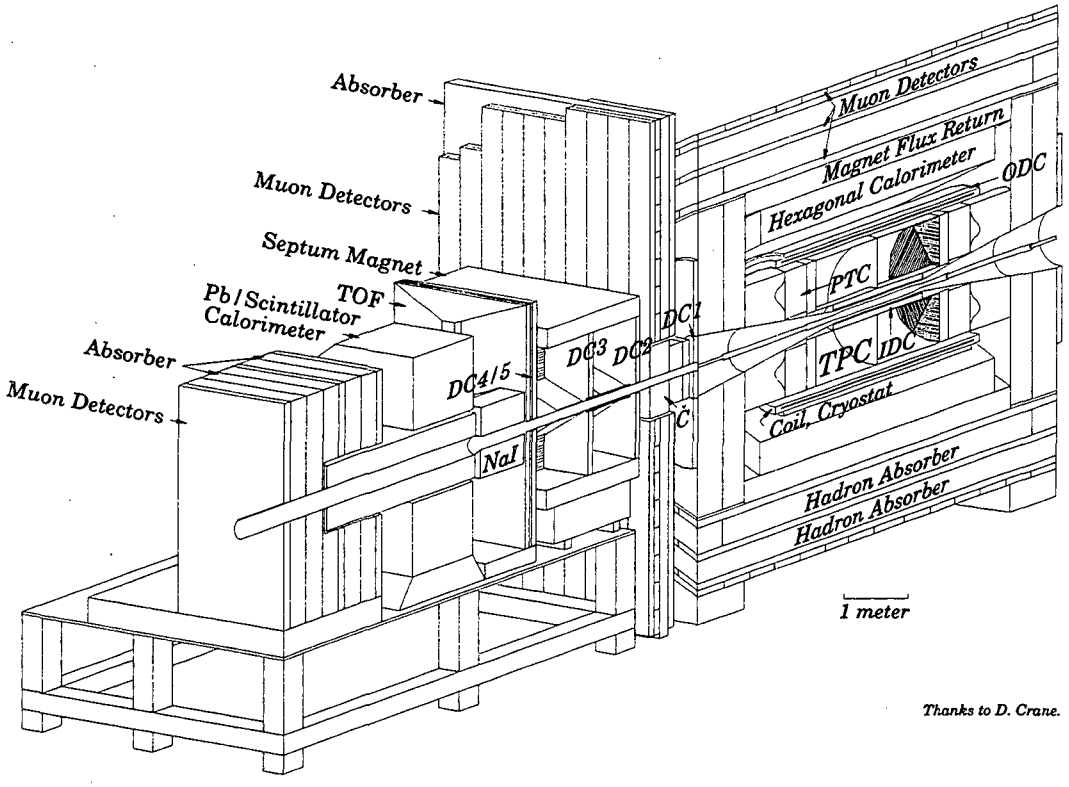


Figure 3.2: A schematic 3-D representation of the TPC/2γ detector, including the forward detectors (From [67]).

A standard coordinate system is used for defining positions. The x direction points horizontally outward from the center of the PEP ring, roughly east, and the y direction points upward. The z axis is the axis of the TPC, at the center of the beam pipe, and the positive z direction points along the electron beam direction, roughly south. The origin of this coordinate system is at the geometric center of the TPC. The interaction point where the electron and positron beams collide is approximately at the origin. The dip angle, λ, is the angle from the plane defined by $z = 0$, the positive direction being in the positive z direction.

This analysis only uses charged particle information from the TPC, so only this part of the detector is described here.

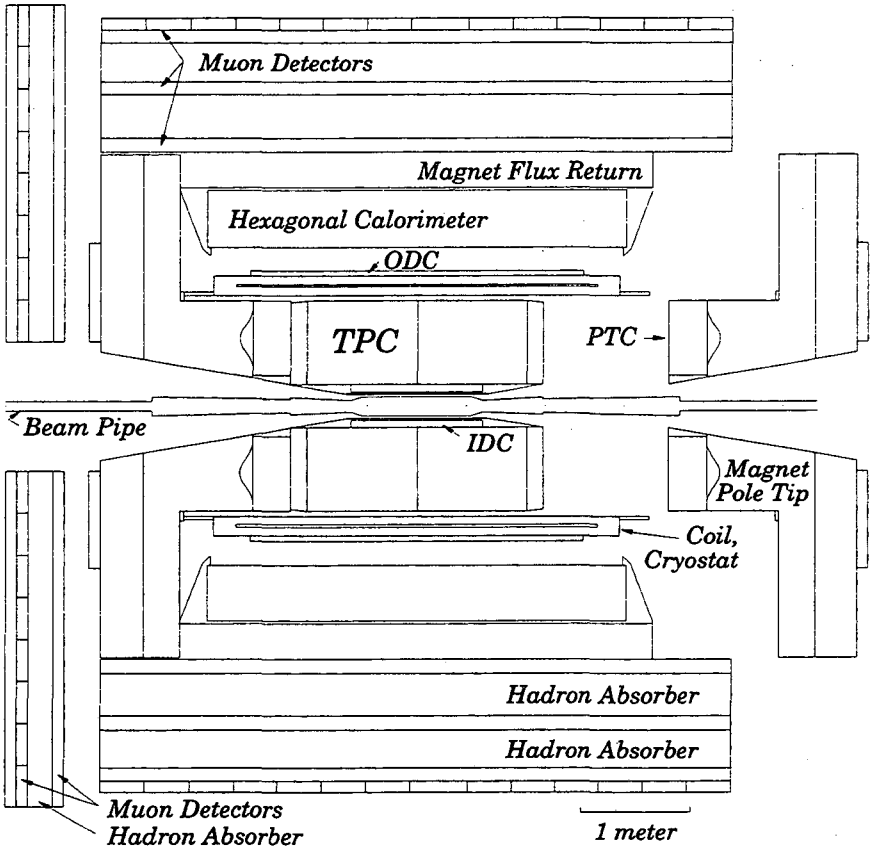


Figure 3.3: A schematic side view of the TPC/2γ detector (From [67]).

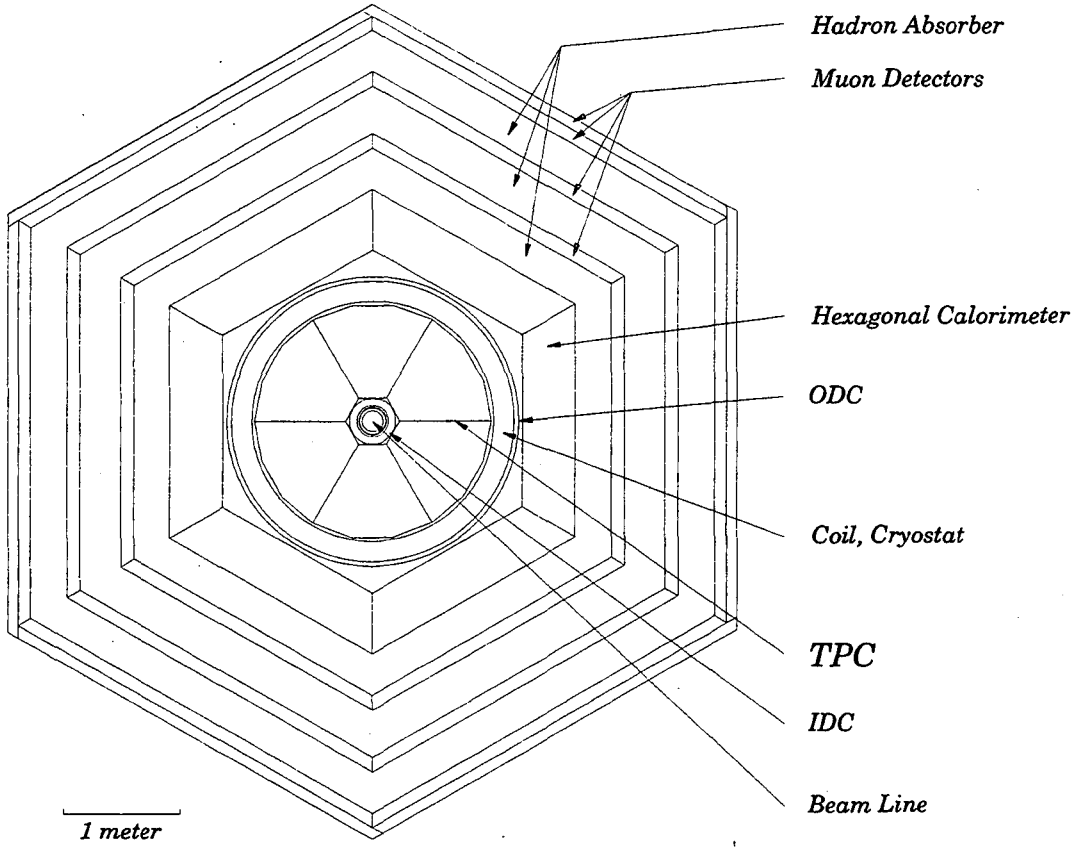


Figure 3.4: A schematic end view of the TPC/2γ detector (From [67]).

3.2 The Time Projection Chamber

The Time Projection Chamber is a drift chamber 2 m long and 1 m in radius, filled with a 80% argon/20% methane mixture at 8.5 atm and immersed in electric and magnetic fields both parallel to the common axis of the TPC and the entire detector. The 13.25 kG magnetic field is generated by the superconducting coil. The electric field is produced by a wire mesh, midway between the ends of the TPC at $z = 0$, held at a voltage of -50 kV or -55 kV with respect to the grounded ends of the TPC. The uniformity of the electric field is produced by metallic equipotential rings, joined by high precision resistors, in the G-10 walls of the TPC.

Each end, or *end cap*, of the TPC is made of six multiwire proportional chambers, or *sectors* (Figure 3.5). Each sector has 183 sense wires spaced 4 mm apart and 4 mm above 15 rows of 7.0 mm by 7.5 mm cathode pads (Figure 3.6). The electric field lines start on the midplane and end on the sense wires, near which there are large electric fields because of the convergence of the field lines. The sense wires are interleaved with field wires that shape the electric field near the sense wires (Figure 3.6). Located 4 mm above this plane of sense and field wires is a grounded grid of wires, which defines the ground seen by the high voltage midplane. At a distance of 8 mm above the grounded grid is another plane of wires, called the gating grid, which keeps positive ions created at the sense wires (discussed below) from reaching the TPC volume. The gating grid acts as an electronic door that is closed except during the brief periods when the trigger electronics [78, 79, 80] have decided to record track information.

As a charged particle traverses the TPC volume, it interacts electromagnetically with gas molecules along its trajectory. This interaction causes the traversing particle to lose energy to these molecules, ionizing them, and leaving an ionization trail, or *track*, along the path that the particle took through the TPC. The ionization electrons drift along the electric field lines from the TPC volume to the sense wires, where they are accelerated so much that they ionize other molecules, creating an avalanche of electrons onto the sense wires, leaving behind positive ions.

Knowing which cathode pads have an electric charge capacitatively induced from the sense wires above them gives information on the projection of a track onto the x-y plane. The z coordinate of a track segment is measured by the arrival time of the segment's ionization electrons. Each pad row contributes one space point in 3 dimensions, for a maximum of 15 space points. These space points are used to reconstruct the trajectory of the particle that created the track. The sense wires are also used to record the amount of ionization per unit track length ($\mathrm{d}E/\mathrm{d}x$), which is used to estimate the velocities of the charged particles that created these tracks.

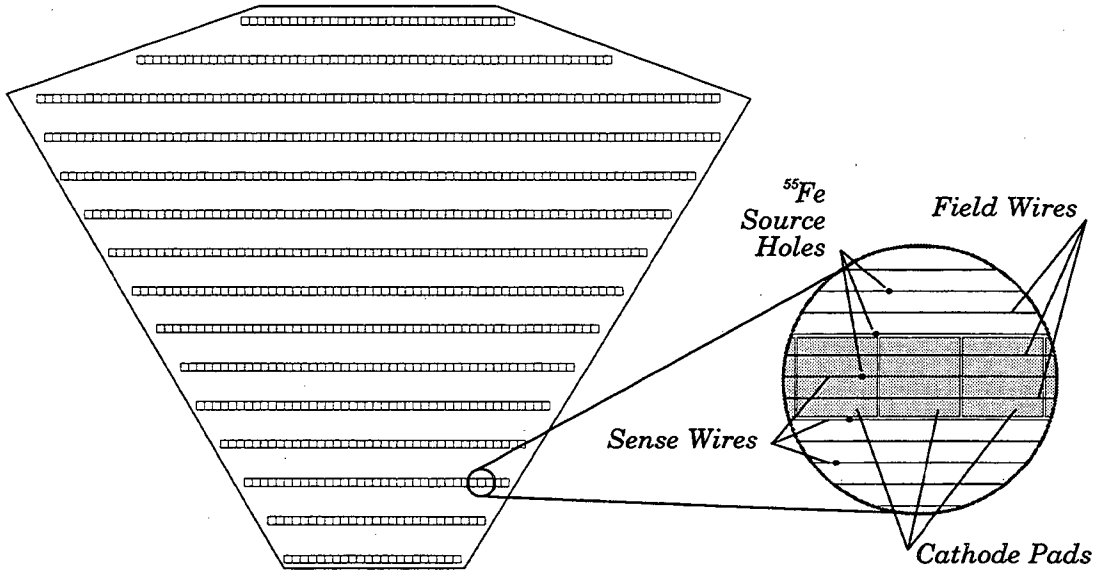


Figure 3.5: A TPC sector (From [67]).

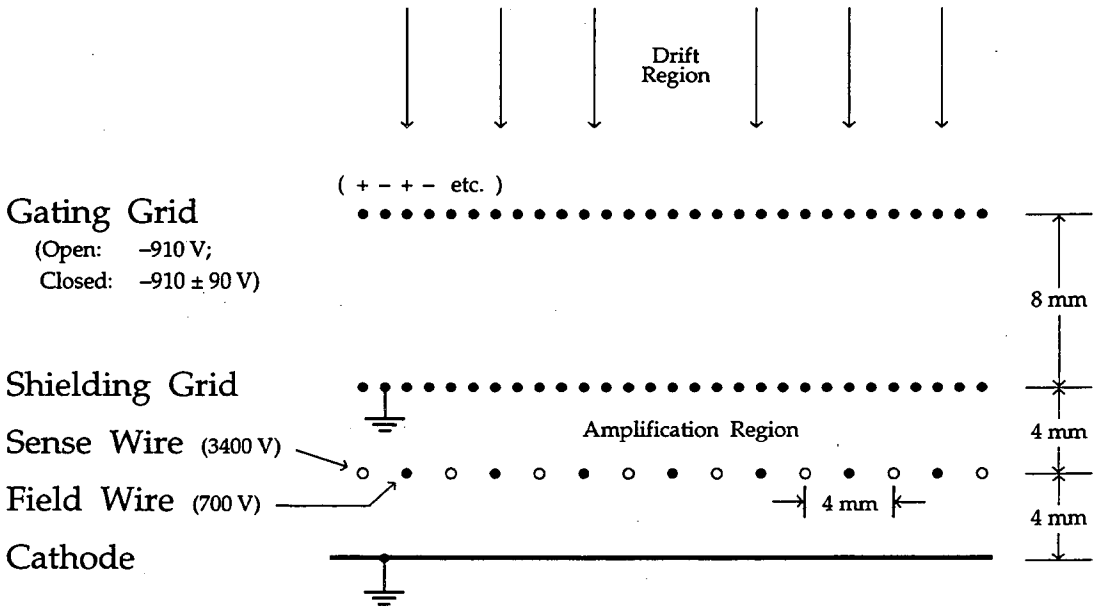


Figure 3.6: The Time Projection Chamber wires (Based on [81]).

Each sense wire and pad in the TPC is connected to a channel in the TPC electronics. For each electronics channel, a preamplifier integrates the collected charge and produces a signal with a fast rise time and a 5 μs decay time. This pulse goes to an amplifier in the electronics house, which generates a roughly Gaussian signal with a 250 ns width that is sampled at 100 ns intervals and stored in a CCD analog shift register. Each CCD bucket is digitized with 9 bit accuracy, and those buckets exceeding a software controlled threshold are read out into the Large Data Buffer and recorded by the VAX 11/782 online computer for data analysis.

3.3 Calibration of the TPC

In order to use $\mathrm{d}E/\mathrm{d}x$, the response of the sense wires to drifting electrons must be calibrated. Before the TPC was assembled, detailed maps were made of how wire gain varied across each sector. The gain was found to vary by about 3% because of non-uniformities in wire diameter and the distance from the wires to the pads. Variations of the gain with time were measured using ^{55}Fe x-rays to produce pulses on the sense wires. Each sector has three ^{55}Fe source rods, at 0, -15, and $+30$ degrees from the sector midline (Figure 3.5).

The dependence of gain upon the sense wire pulse amplitude was measured by pulsing the voltage on the shielding grid with 11 different amplitudes; this induces pulses on the sense wires. The coupling between the shielding grid and the sense wires is not known well enough to normalize the gain curve, so the ^{55}Fe data and minimum ionizing pions were used to obtain a normalization.

Chapter 4

Particle Identification Using the TPC

In this chapter, we describe how particles are identified by the TPC using simultaneous measurements of momentum and $\mathrm{d}E/\mathrm{d}x$. First, we show how momentum is measured using the TPC. Then, we describe the dependence of the mean of $\mathrm{d}E/\mathrm{d}x$ of a charged particle upon the particle's speed. Finally, we describe the resolution of $\mathrm{d}E/\mathrm{d}x$ measurements, and the parameterization of this resolution.

4.1 The Measurement of Momentum

The momentum of a charged particle is obtained from the curvature of the particle's reconstructed trajectory in the magnetic field [82]. If the radius of curvature of a particle trajectory is R, then the momentum of the particle in the plane perpendicular to the TPC axis is

$$p_\perp = \frac{BR}{3335} , \tag{4.1}$$

where R is in cm, B is the magnetic field strength in kG, and $p_\perp$ is in GeV/c. The reconstruction of the track provides the angle λ, allowing us to find the track momentum

$$p = \frac{p_\perp}{\cos\lambda} . \tag{4.2}$$

The momentum has an average resolution of

$$\left(\frac{\Delta p}{p}\right)^2 = (0.015)^2 + (0.007\, p)^2 , \tag{4.3}$$

where the average is over track length and λ. This resolution is caused by measurement errors and Coulomb scattering of the track.

4.2 The Theory of $\mathrm{d}E/\mathrm{d}x$ Energy Loss

The deposited energy per unit trail length ($\mathrm{d}E/\mathrm{d}x$) is a function only of the particle's speed. Measuring $\mathrm{d}E/\mathrm{d}x$ and momentum simultaneously therefore allows us to deduce a particle's mass, thus identifying the particle. This ability to identify particles is a crucial part of the study presented in Chapter 8.

$\mathrm{d}E/\mathrm{d}x$ is reflected in the amount of charge collected on the sense wires. Each wire makes a measurement of $\mathrm{d}E/\mathrm{d}x$, since each wire collects a sample of $\mathrm{d}E/\mathrm{d}x$ ionization electrons from a 4 mm thick slice of gas, which corresponds to an average of 5 mm of track length.

Electrons are ionized from gas molecules in two ways, depending on the amount of energy transferred to the gas molecule. For small energy transfers, the ionization cross-section is peaked at the electron binding energy, and since small energy transfers are most probable, this is the most important mechanism for energy transfer. For large energy transfers, the gas electrons are basically free and the process is described by Rutherford scattering. Rutherford scattering is relatively rare, but contributes a lot of ionization energy through these rare scatters (the so-called Landau tail in Figure 4.1), so the total ionization has large statistical fluctuations. To reduce these statistical fluctuations, the average of the smallest 65% of the individual wire $\mathrm{d}E/\mathrm{d}x$ measurements is used as the measure of $\mathrm{d}E/\mathrm{d}x$. This is called the *truncated mean*, or the "$\mathrm{d}E/\mathrm{d}x$" of a track.

The dependence of $\mathrm{d}E/\mathrm{d}x$ on $\beta\gamma$ is shown in Figure 4.2 [83]. For slow-moving particles, $\mathrm{d}E/\mathrm{d}x \propto 1/\beta^2$, and $\mathrm{d}E/\mathrm{d}x$ drops sharply with increasing β. This behavior is due to the fact that when the particle moves slowly, it spends more time near gas atoms and is more likely to ionize them, and this effect diminishes in strength as β increases. $\mathrm{d}E/\mathrm{d}x$ reaches a minimum at around $\beta\gamma = 3$, which is called the *minimum ionizing* region. $\beta\gamma$ between 3 and about 10^3 is called the *relativistic rise* region, where relativity enhances the particle's transverse electric field and its ability to ionize the medium, causing $\mathrm{d}E/\mathrm{d}x$ to rise slowly with $\beta\gamma$. In the Fermi plateau region, above $\beta\gamma \approx 10^3$, the curve flattens out as the medium polarizes in response to the transverse field, cutting off any further increase in ionizing power [84, 85].

$\beta\gamma$ is not directly measured, however, momentum is. $p = \beta\gamma m$ implies $\log(p) = \log(\beta\gamma) + \log(m)$, so when $\mathrm{d}E/\mathrm{d}x$ is plotted as a function of $\log(p)$ for different particle species, the curves for different species are simple translations, with respect to $\log(p)$, of the same curve that traces how $\mathrm{d}E/\mathrm{d}x$ depends upon $\log(\beta\gamma)$. The theoretical dependence of $\mathrm{d}E/\mathrm{d}x$ upon $\beta\gamma$ in the TPC has been calculated elsewhere, and fitted to data [25]. The fit is excellent.

The measured mean $\mathrm{d}E/\mathrm{d}x$ also varies with time, dip angle, and TPC sector. These variations are corrected for by calculating how the average $\mathrm{d}E/\mathrm{d}x$ for minimum ionizing pions depends on these quantities [25, 86].

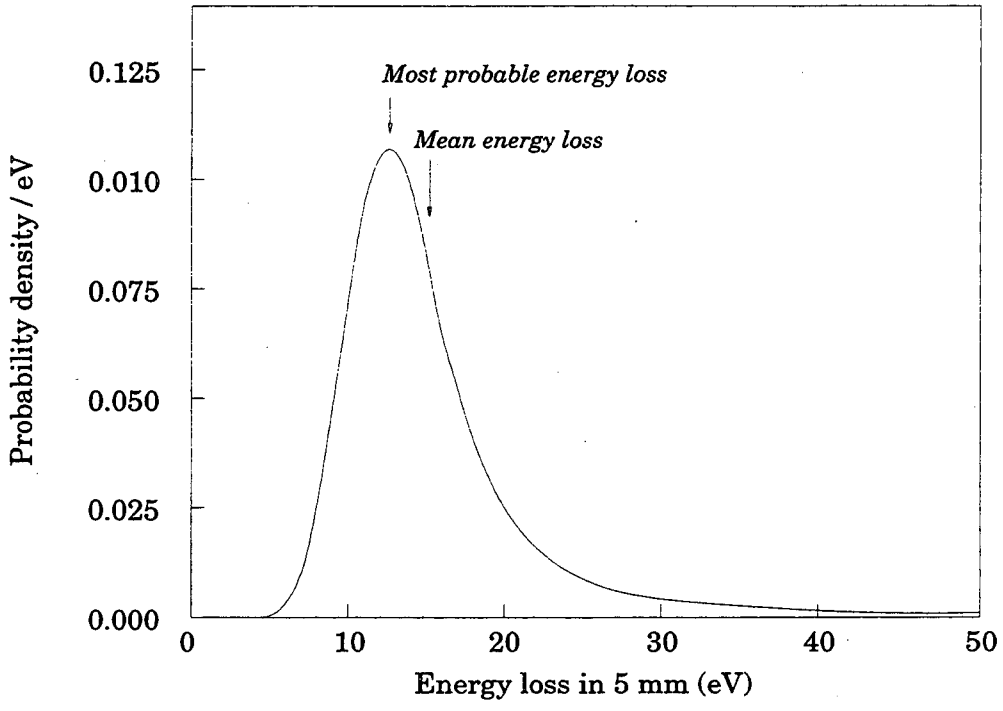


Figure 4.1: Predicted distribution of $\mathrm{d}E/\mathrm{d}x$ energy loss for a 5 mm track sample (From [67]).

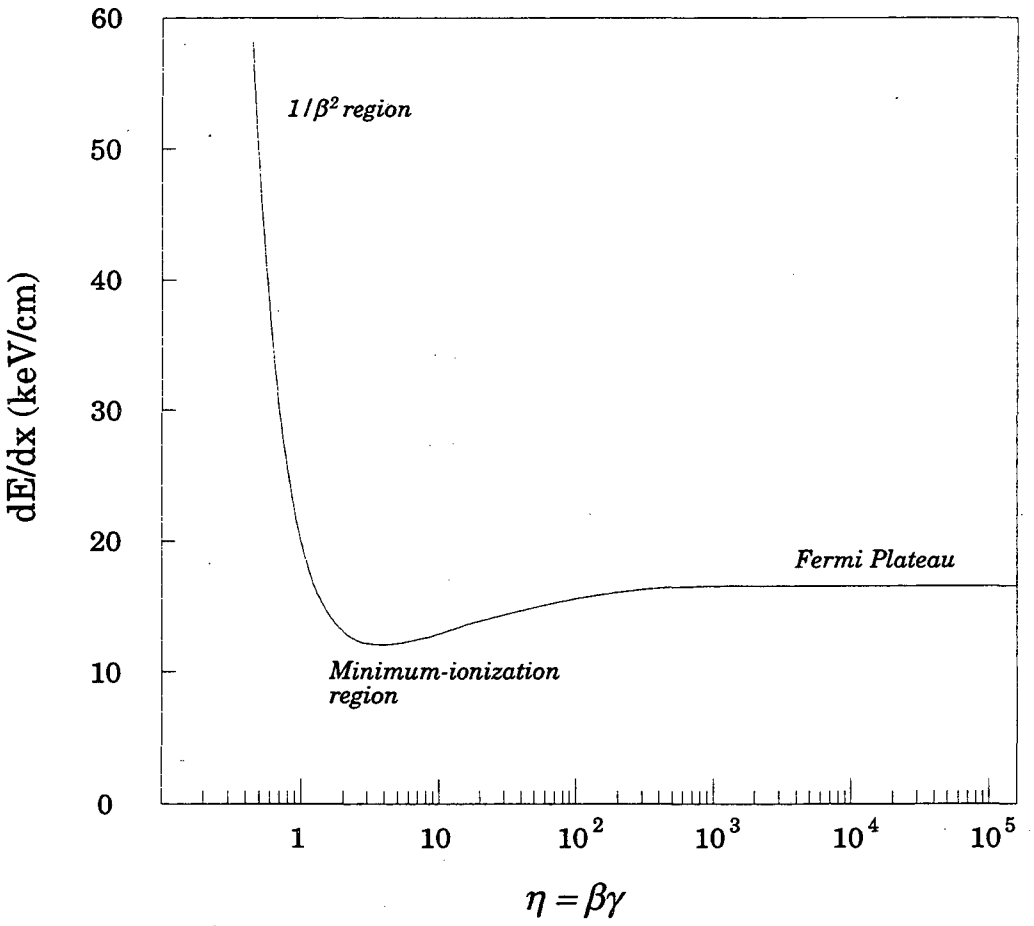


Figure 4.2: Dependence of $\langle \mathrm{d}E/\mathrm{d}x \rangle$ on $\eta = \beta\gamma$ (From [67]).

4.3 dE/dx Resolution Parameterization

Each track has only a finite number of dE/dx samples, therefore experimentally measured values of dE/dx have a finite resolution and form bands about the theoretical curves. Figure 4.3 shows that these bands sometimes are not well separated, particularly at high momentum and near the points where the theoretical curves cross. In these regions, tracks can not be unambiguously associated with only one dE/dx curve and can not be assigned a unique particle identification. We can still statistically determine the number of tracks with each particle identification, because we know the behavior of the dE/dx resolution, which we now describe.

The resolution has a Gaussian distribution out to at least 3 standard deviations [25]. For a sample of minimum ionizing pions from the High-Field data set with at least 120 wire hits, the resolution is 3.4%[1] [25]. The resolution varies with time, the number of wires with dE/dx samples (N), and $|\sin\lambda|$.

We now describe the method used previously to find the parameterization for how dE/dx depends on N and $|\sin\lambda|$ [25]. To take into account time variation in the resolution, the time in which the TPC collected the High-Field data set was divided into 10 intervals[2] in such a way that the samples of minimum ionizing pions within these intervals are approximately the same size. Each sample was analyzed separately. For each sample, the standard deviation σ of the quantity

$$R = \frac{\text{measured Trmean}(p_{tpc})}{\text{expected Trmean}(p_{tpc})} \tag{4.4}$$

was computed and histogrammed as a function of N (p_{tpc} is the momentum in the TPC). It was found that the parameterization $\sqrt{A/N + B}$, with A and B free parameters, fits this plot pretty well. On average, A is 0.105 and B is 3.6×10^{-4}. $\sqrt{A/N}$ is the dependence of σ on the number of measurements, while the B term could be an intrinsic resolution added in quadrature to $\sqrt{A/N}$. What was then done was to compute and histogram the standard deviation of $(R-1)/\sqrt{A/N+B}$ as a function of $|\sin\lambda|$, in several histograms each with a different range of N. It was found that each histogram was well fit by a line, but the y-intercept of the fitted line varied with N. The end result was the parameterization

$$\sigma(N,\lambda) = \sqrt{A/N + B}\ (C + DN + E|\sin\lambda|)\ . \tag{4.5}$$

C was found to be 1.17, D was found to be -7.2×10^{-4}, and E was found to be -0.28.

[1]For the Vertex Chamber data set, which is not used in this analysis, the corresponding sample of pions has a dE/dx resolution of 3.1% [25, 86], the improvement due to better monitoring of the TPC.

[2]The Vertex Chamber data set was divided into 5 intervals.

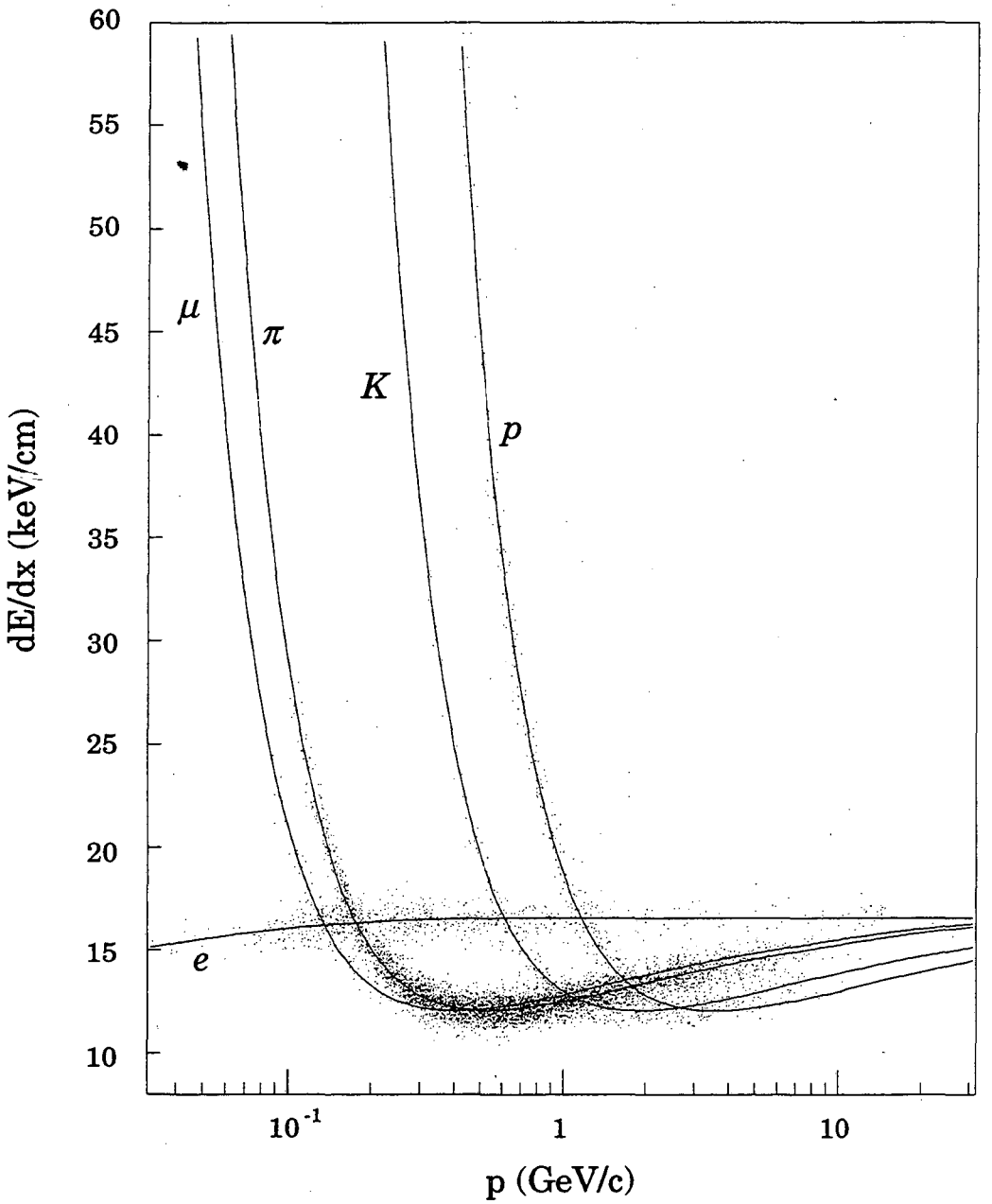


Figure 4.3: Distribution of a sample of tracks in dE/dx as a function of $\log(p)$, with the predicted curves for e, μ, π, K, and p (From [67]).

This previously used method of parameterization contains a subtle but important oversight. The scatterplot of the High-Field data set pion tracks in the N-$|\sin\lambda|$ plane at large $|\sin\lambda|$ (Figure 4.4) has a diagonal concentration of tracks[3], therefore the distribution of tracks in N is not independent of the distribution of tracks in $|\sin\lambda|$. As a result, the standard deviation as a function of N only, with the dependence on $|\sin\lambda|$ projected out, does *not* find how the resolution depends on N as desired. Instead, the standard deviation with its dependence on $|\sin\lambda|$ projected out is influenced by the variation of the resolution *along* the concentration of tracks at large $|\sin\lambda|$. The end result is that the best fit of a Gaussian to the distribution of $(R-1)/\sigma(N,\lambda)$ for all pions with 40 or more wire hits has a width of 1.1053 ± 0.0038, which is more than 27 standard deviations larger than the expected value of 1, so the particle ID χ^2 values are 10% too big, on average.

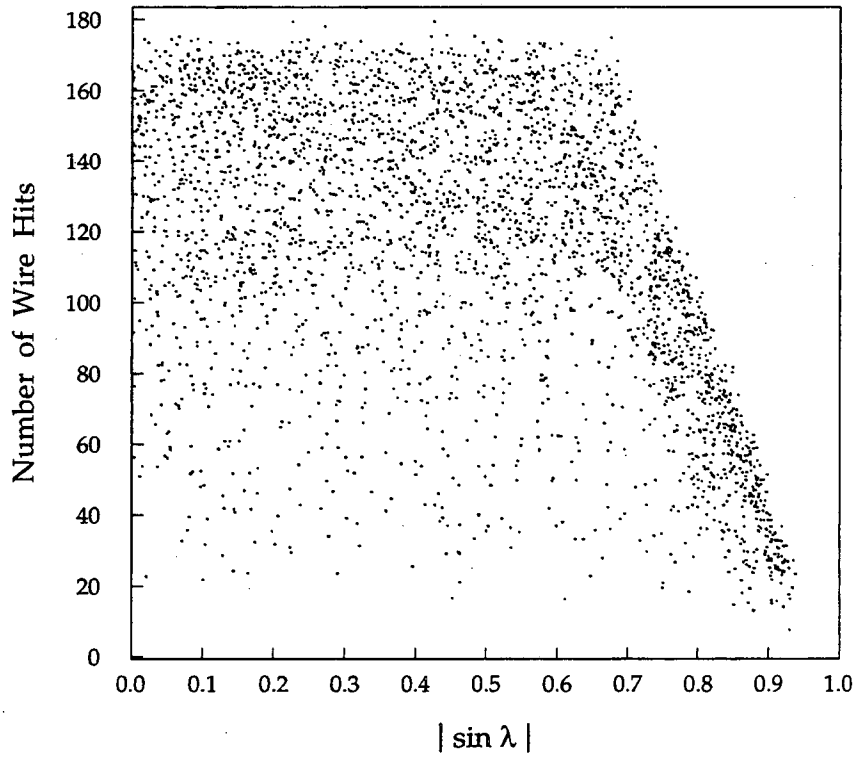


Figure 4.4: Scatter plot of N, the number of wire hits, as a function of $|\sin\lambda|$ for minimum ionizing pions.

To fix this problem, the old parameterization of Equation 4.5 is fit in N and

[3]This concentration is produced by the fact that for $|\lambda|$ greater than 42°, tracks coming from the Interaction Point can not leave ionization electrons on all of the 183 sense wires, since the ionization of that portion of a track with $|z| > 0.9$ m is absorbed by the gating grid. As a result, these large-dip tracks can not deposit charge on all of the 183 wires, only a subset, but most tracks will deposit charge on most of the wires accessible to the ionization electrons.

$|\sin\lambda|$ simultaneously. In this fit, N is restricted to be greater than or equal to 40, since the resolution of the data for N less than 40 is systematically lower than the fit. The N-$|\sin\lambda|$ plane was divided into rectangular bins, and $\sigma(R)$ in each bin was obtained with a maximum likelihood fit of a Gaussian to the distribution of R in that bin. Then $\sigma(N,\lambda)$ was calculated at the "center of gravity" of the data in each bin and fit to the $\sigma(R)$'s, using a χ^2 fit. The standard deviation of the $\sigma(R)$ in each bin is $\sigma(\sigma(R)) = \sigma(R)/\sqrt{2n}$, where n is the number of tracks in the bin. This expression for $\sigma(\sigma(R))$ is biased, however, since a small value of $\sigma(R)$ makes the standard deviation of $\sigma(\sigma(R))$ small also. Instead, we use $\sigma(\sigma(R)) = \sigma(N,\lambda)/\sqrt{2n}$.

When the fit of $\sigma(N,\lambda)$ is done in this manner, B and D in Equation 4.5 are zero, within statistics, for both the High-Field data set and the unused Vertex Chamber data set. The fit is good. B and D were set to zero, and the fit redone with the resolution parameterized by

$$\sigma(N,\lambda) = \frac{C + E|\sin\lambda|}{\sqrt{N}}\ , \tag{4.6}$$

which is much simpler than the original form. On average, C was found to be 0.474 and E was found to be -0.146 for the High-Field data set[4].

The new parameterization we use does not suffer from the problems of the old parameterization: when $(R-1)/\sigma(N,\lambda)$ is plotted for all pions with 40 or more wire hits, the best fit Gaussian has a width of 0.9989 ± 0.0034 (Figure 4.5), which is statistically compatible with the expected value of 1.

[4]For the Vertex Chamber data, on average, C was found to be 0.440 and E was found to be -0.100.

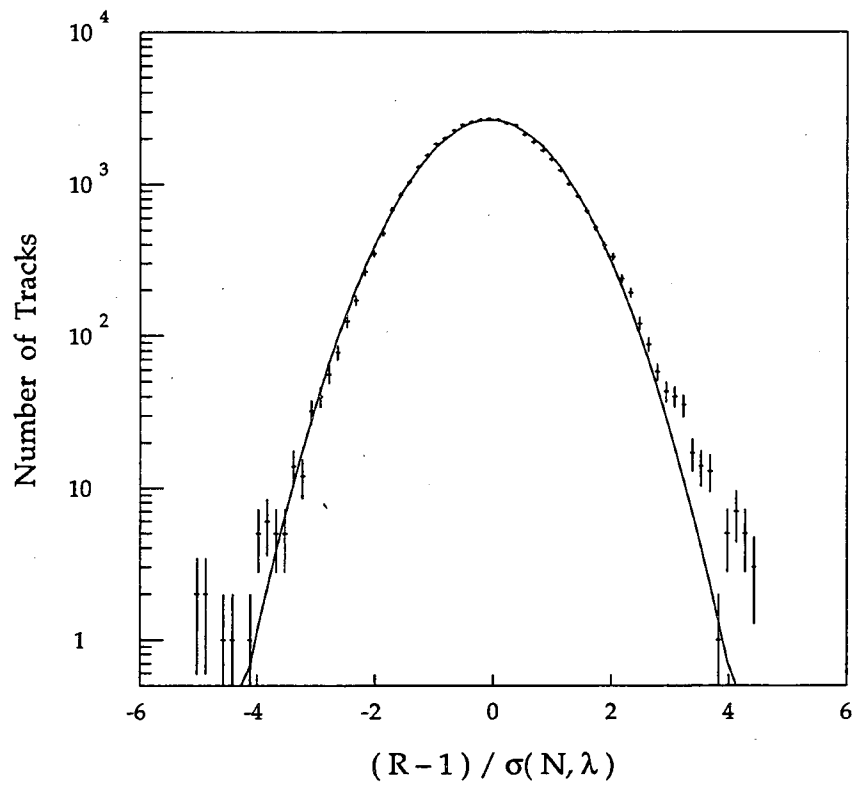


Figure 4.5: Number of minimum ionizing pions (with $N \geq 40$) as a function of $(R-1)/\sigma(N,\lambda)$ (points with statistical error bars), with a Gaussian fit (solid).

Chapter 5

Event Reconstruction, Selection, and Simulation

In this chapter, we describe the more basic tools we use to do this analysis on the collected data. First, we outline the processing of raw data into reconstructed events. Then, we describe the selection criteria for removing backgrounds to the signal process $e^+e^- \rightarrow q\bar{q} \rightarrow hadrons$. Finally, we describe the Monte Carlo package used to simulate the experimental data and the tuning of this package so that it properly simulates experimental data.

5.1 Event Reconstruction

The TPC electronics, described in Chapter 3, produce a large amount of raw data that must be reconstructed into the events that produced the raw data. The reconstruction of events is done by a set of programs [87] that processes ADC counts into space points, associates space points and wire hits with charged particle tracks, calculates the dE/dx of tracks from wire hits, makes particle identification assignments, fits tracks to a common event vertex, and removes cosmic ray tracks from the data. The end product of the event reconstruction are Data Summary Tapes (DSTs) that list the fitted tracks and their characteristics. Raw data is excluded from the DSTs.

5.2 Event Selection

When electrons and positrons collide in the PEP ring, they interact in a variety of ways, including annihilation into $\mu^+\mu^-$, $\tau^+\tau^-$, or $q\bar{q}$ pairs, Bhabha scattering ($e^+e^- \rightarrow e^+e^-$), and 2-photon processes ($e^+e^- \rightarrow e^+e^-\gamma\gamma$ and $\gamma\gamma \rightarrow X$, where X

= anything). This analysis is only concerned with the process $e^+e^- \rightarrow q\bar{q}$, so we need to filter out, as much as is practical, events due to the other processes.

The standard Good Hadronic Event Selection used in all TPC analyses of the process $e^+e^- \rightarrow q\bar{q}$ is done by a program called Label_Hadron_v2, which uses as its input the DSTs produced at the end of the data reconstruction. The selection criteria take advantage of the typical characteristics of hadronic events: large multiplicities, large energies, and low boost of the final state's center of mass. Only charged tracks are used in this selection, and they must have these characteristics:

1. The dip angle $|\lambda|$ of the track must satisfy the constraint $|\lambda| < 60°$, to ensure that the track enters the TPC fiducial volume.

2. Either $dC < 0.30$ GeV^{-1} or $dC/C < 0.30$, where C is the track's measured curvature, and dC is the error on the measured curvature. This ensures that the track is well measured.

3. The track momentum in the TPC volume must be > 100 MeV/c, and the track momentum extrapolated to the interaction point must be > 120 MeV/c. This ensures that the track has enough energy to have come from the event vertex.

4. The track must come from within 6 cm of the nominal vertex in the x-y plane and from within 10 cm in z, where the nominal vertex is the average vertex position for a period of roughly a few hours. This rejects tracks coming from cosmic rays or from interactions of particles with atomic nuclei in the beampipe.

Tracks that satisfy these criteria are called *good tracks.*

To pass the Good Hadronic Event Selection, an event must satisfy these criteria:

1. There must be five or more good tracks that are not electrons. The electron identification is done either by dE/dx or by an algorithm that identifies pairs of tracks as coming from photon conversions. This rejects showering Bhabha events and low multiplicity events, including τ_{1+3} events[1].

2. The total energy of the good tracks, $E_{charged}$, must be > 7.25 GeV. This rejects 2-photon and beam-gas events.

3. The reconstructed event vertex must be within 2 cm in the x-y plane and within 3.5 cm in z of the nominal vertex. This rejects beam-gas events.

[1] "1+3" signifies that one τ lepton decays into a final state with 1 charged particle and the other decays into a final state with 3 charged particles. 97.8% of τ events have 4 or fewer charged particles in the final state [88].

4. At least one hemisphere of the event, as defined by the plane perpendicular to the sphericity axis, must contain either

   (a) four or more good non-electron tracks or

   (b) good tracks with an invariant mass greater than 2 GeV/c^2.

   This criterion rejects τ_{1+3} events.

5. To reject 2-photon, beam-gas, and energetic ISR events, all of which usually have a large momentum imbalance in the z direction, the sum of the z-component of the momentum for all tracks, $\sum p_z$, must satisfy $|\sum p_z| < 0.4 E_{charged}$.

6. At least half of the tracks must be good tracks, to eliminate problematic events.

Monte Carlo simulation of the process $e^+e^- \rightarrow q\bar{q}$ (discussed in the next section) shows that this set of selection criteria rejects approximately 32% of the $e^+e^- \rightarrow q\bar{q}$ events. For the High-Field data set, 25189 events satisfy these selection criteria.

In this analysis, we use two additional selection criteria to ensure that the bulk of each event enters the fiducial volume of the TPC:

1. The sphericity axis of the event must be more than 45° from the beam direction.

2. The sphericity of the event must be less than 0.5, to ensure that the direction of the event axis is meaningful.

19529 events satisfy these additional selection criteria, as well as the Good Hadronic Event Selection.

Previous Monte Carlo studies have estimated that, of the events that pass the Good Hadronic Event Selection, (0.4 ± 0.1)% are τ events and (0.5 ± 0.1)% are two-photon events [81]. To remove these τ and 2-photon events, we use the additional 'τ-2γ' selection on events containing fewer than 9 good hadrons:

1. To remove τ events, we discard events with an event thrust greater than 0.97.

2. To remove 2-photon events, we discard events with $E_{charged}$ less than 9.5 GeV. A study [89] shows that essentially all 2γ events have charged energy less than 9.5 GeV.

The motivation for using these two additional selection criteria is discussed in Section 5.3.3.4. We estimate that these two criteria remove 94% of all τ events and 11% of the hadronic events that pass the Good Hadronic Event Selection. We have no estimate of the fraction of the 2-photon events that pass the good hadronic selection and are cut by these two selection criteria. 17943 events pass all of the selection criteria discussed in this section. The effect of all of the selection criteria upon the measured bottom event fraction is discussed in Section 7.4.3.

5.3 Event Simulation

5.3.1 The Event Simulation Software Package

In doing any analysis, it is important to simulate the physics processes that produce the observed events and the effects of the detector on the events collected. We use three programs for this simulation. Jetset 7.2 is used to generate simulated e^+e^- $\rightarrow$ *hadrons* events, except that we use the spring 1992 version of the CLEO Monte Carlo (described below) to simulate the weak decays of charm hadrons and bottom mesons.

Jetset 7.2 implements the Lund string fragmentation model, as described in Section 2.4.2, with parton showers. The Lund Symmetric Fragmentation Function (LSFF) does a fine job of reproducing the distributions of simple variables (such as sphericity, thrust major, charged multiplicity, and track momentum) for all hadronic events. However, the LSFF is unable to reproduce simultaneously the distributions of these variables, the average momentum of charm hadrons, and the average momentum of bottom hadrons [46]. This is unacceptable, since our analysis requires an accurate simulation of bottom, charm, and light quark events. Our solution is to use the LSFF only for light quark events, and to use the Peterson function, described in Section 2.4.4, for charm and bottom events. The Lund Monte Carlo uses a number of free parameters to determine how it hadronizes events, so these parameters must be tuned before Jetset is able to reproduce important characteristics of experimental data. The parameters that govern hadron flavor production, including production of bottom and charm hadrons, have already been tuned to data. Table 5.1 lists the values of those flavor parameters that are different from the default values [90]. The tuning of the parameters that govern event shape is discussed in Section 5.3.3.

The portion of the CLEO Monte Carlo used to simulate the decays of charm and bottom hadrons is an up-to-date *decay table*, which is a list of exclusive and inclusive decay modes, along with their branching ratios and decay matrix elements. This list instructs the CLEO Monte Carlo how to simulate charm and bottom hadron decays. The list contains all of the known exclusive decays, along with inclusive decay modes (e.g. $b \rightarrow c\bar{u}d$) that have been tuned so that the Monte

Lund parameter[3]	Tuned value	Default value
$P(qq)/P(q)$	0.089	0.10
$P(s)/P(u)$	0.282	0.30
$[P(su)/P(du)]/[P(s)/P(u)]$	0.74	0.4
$1/3\ P(ud1)/P(ud0)$	0.03	0.05
$P(BMB)/P(BMB+BB)$	0.9	0.5
$P(vector)/P(all)$ for u,d	0.46	0.5
$P(vector)/P(all)$ for s	0.46	0.6
$P(vector)/P(all)$ for c,b	0.59	0.75

Table 5.1: Tuned values of the Lund flavor parameters in Jetset 7.2.

Carlo reproduces experimentally measured properties of charm and bottom decays. The CLEO Monte Carlo has been used for a long time and is known to reproduce the properties of bottom meson decays. We have verified that the CLEO Monte Carlo reproduces experimentally measured multiplicities and momentum spectra for charged pions, kaons, and protons from bottom meson decays, while Jetset does not.

TPCLUND6, the standard fast TPC Monte Carlo, is used to simulate the TPC detector's response to the events generated by Jetset 7.2 and the CLEO Monte Carlo. The output of these three programs are DSTs with the same format as experimental data. These output DSTs are subjected to the same hadronic event selection as experimental data. TPCLUND has been used for many years and is known to accurately simulate TPC track measurements and the acceptance of the TPC[2].

5.3.2 Tuning the Peterson Parameterization

As mentioned in Section 2.4.4, the Peterson function has one free parameter, ϵ. Let $x_E = E/E_{\text{beam}}$, where E is the energy of the charm or bottom hadron in question, and let $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$ be the average x_E for primary bottom and primary charm hadrons, respectively. ϵ for charm events, ϵ_c, must be adjusted so that the Monte Carlo generated $\langle x_E \rangle_c$ matches the experimentally observed $\langle x_E \rangle_c$. Likewise, ϵ

[2]The previous version of TPCLUND, version 5, was mated to Jetset 5.3 and 6.3. The structure of the programming of Jetset, especially the common blocks, is substantially different from that of Jetset 7.2. Rather than restructure TPCLUND to follow this change, the mating of TPCLUND to Jetset 6.3 was retained and the features of Jetset 7.2 that differ from Jetset 6.3 were copied into our now custom version of Jetset 6.3. As a result, when we refer to specific locations in the Jetset common blocks, we use the Jetset 6.3 locations.

for bottom events, ϵ_b, must be adjusted so that the Monte Carlo generated $\langle x_E \rangle_b$ matches the experimentally observed $\langle x_E \rangle_b$.

Numerical values of $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$ have been obtained from the plots in Reference [49] and are listed in Table 5.2. A 1-σ error of ± 0.002 is assigned to these numbers from the uncertainty in the method used to read these values from the plots. We also assign an error of ± 0.03 [92] to these numbers to reflect the statistical and systematic uncertainties on $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$ in Reference [49].

Quantity	Charm	Bottom
$\langle x_E \rangle + 2\sigma$ error	0.586	0.775
$\langle x_E \rangle$ central value	0.526	0.715
$\langle x_E \rangle - 2\sigma$ error	0.466	0.655

Table 5.2: $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$ from Reference [49].

Quantity	Charm	Bottom
ϵ for $\langle x_E \rangle$ upper limit	0.0153	0.0072
ϵ for $\langle x_E \rangle$ central value	0.072	0.039
ϵ for $\langle x_E \rangle$ lower limit	0.258	0.148

Table 5.3: ϵ_b and ϵ_c for Jetset 7.2.

In Table 5.3 are listed the values of ϵ_b and ϵ_c in Jetset 7.2 that reproduce these values of $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$.

5.3.3 Tuning the Jetset Event Shape Parameters

The Lund parameters to be tuned govern the shape and momentum structure of Monte Carlo-generated events. They are:

[3] $P(qq)/P(q)$ is the diquark to quark probability ratio in breaking the string. $P(s)/P(u)$ is the ratio of strange quark to up (or down) quark production. $[P(su)/P(du)]/[P(s)/P(u)]$ is the probability of creating an su-diquark versus a du-diquark, relative to the ratio $P(s)/P(u)$. $P(ud1)$ and $P(ud0)$ are the probabilities for the creation of a spin-1 and a spin-0 diquark, respectively. $P(BMB)/P(BMB + BB)$ is the 'popcorn' baryon production [91] probability, where the meson M is created between two baryons B in string fragmentation. $P(vector)/P(all)$ is the fraction of mesons produced that are vector.

1. the a and b parameters in the Lund Symmetric Fragmentation Function (Section 2.4.2),
2. Λ_{QCD}, the QCD scale parameter described in Section 2.3,
3. σ_q, the width of the Gaussian p_t distribution (Section 2.4.2), and
4. the parton shower virtuality cut-off Q_0 described in Section 2.3.2.

The vector to pseudoscalar ratio r also affects event shape, but it is fixed at the value determined by vector meson production measurements.

The Peterson parameters tuned in the last section are weakly correlated with Λ_{QCD} and Q_0. This implies that the optimization of the 5 Lund parameters listed above throws off the optimization of the Peterson epsilons. In principle, therefore, it is necessary to tune the Lund parameters and the Peterson epsilons alternately until they all converge. It turns out that the tuning in this section changes $\langle x_E \rangle_b$ and $\langle x_E \rangle_c$ by much less than their errors, so the Peterson epsilons need not be retuned.

The procedure used here for tuning the Lund model parameters in Jetset 7.2 largely follows the procedure used to tune Jetset 5.2 [78], which produces distributions of a set of variables in data and in Monte Carlo and minimizes the differences between the two sets of distributions. We now describe the procedure used to tune Jetset 5.2. After that, we describe the procedure we used to tune Jetset 7.2.

5.3.3.1 The Kinematic Variables Used for Tuning.

We first describe the variables used to tune Jetset 5.2. These variables are listed in Table 5.4. Q_1 and Q_2 are the smallest and next smallest eigenvalues of the sphericity tensor[3]. L_1 and L_2 are the thrust minor and thrust major, respectively[4]. $\langle p_{\perp in} \rangle$ and $(p_i)_{\perp in}$ are the average momentum per event and the momentum per

[3]The sphericity tensor is given by

$$S_{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i p_i^2} , \tag{5.1}$$

where p_i^α is the α-component of the momentum for track i.

[4]The thrust major and thrust minor are defined as follows. First, define

$$T = \frac{\sum_i |\vec{n} \cdot \vec{p_i}|}{\sum_i p_i} , \tag{5.2}$$

where $\vec{n}$ is a unit vector. The thrust L_3 is the maximum of T, and the thrust axis is the vector $\vec{n} = \vec{n}_1$ that maximizes T. The thrust major axis $\vec{n} = \vec{n}_2$ is the vector perpendicular to $\vec{n}_1$ that maximizes T, and this maximum of T is the thrust major L_2. Finally, the thrust minor axis $\vec{n}_3$ is perpendicular to $\vec{n}_1$ and $\vec{n}_2$, and the corresponding value of T, L_1, is the thrust minor.

track in the event plane (as defined by the sphericity tensor) and perpendicular to the sphericity axis. $\langle p_{out} \rangle$ and $(p_i)_{out}$ are the average momentum per event and the momentum per track out of the event plane (as defined by the sphericity tensor). $x_p = p/p_{\text{beam}}$, where p is the momentum of a track. ΔM^2_{jet} is the difference between the squared masses of the the two jets in the event, as defined by the plane perpendicular to the sphericity axis. E_{vis} is the visible charged energy of the event.

Set 1	Set 2	Set 3
Q_2	x_p	Q_1
L_2	Charged Multiplicity	L_1
$\langle p_{\perp in} \rangle$		$\langle p_{\perp out} \rangle$
$(p_i)_{\perp in}$		$(p_i)_{\perp out}$
$\Delta M^2_{jet}/E^2_{vis}$		

Table 5.4: Variables used for tuning.

The quantities in Set 1 are most sensitive to hard gluon radiation, so they are sensitive to Λ_{QCD}, which controls the rate of hard gluon radiation. The quantities in Set 2 reflect the hardness of the fragmentation, and thus are sensitive to the Lund a and b parameters. The tuned values of a and b are highly correlated, so b was fixed and a was tuned. The quantities in Set 3 measure the thickness of the event out of the event plane, and thus reflects σ_q. We use the same variables to tune Jetset 7.2.

5.3.3.2 The Old Tuning Methods.

The method for tuning Jetset 5.2 used two sets of histograms: one set was the distribution of the number of events or tracks as a function of these tuning variables for experimental data, and the other set the corresponding distributions for Monte Carlo. These two sets were compared to each other using two different methods.

In the first method, for coarse tuning, we calculate a χ^2 to compare the experimental and Jetset/TPCLUND distributions. Let D_{ij} be the number of experimental data entries in bin i of distribution j, and let $M_{ij}(\vec{p})$ be the number of entries in bin i of distribution j from our Monte Carlo with parameter set $\vec{p}$. After normalizing the data and Monte Carlo histograms to the same number of entries,

$$\chi^2 = \sum_{j=\text{distribution}} \sum_{i=\text{bin}} \frac{(D_{ij} - M_{ij}(\vec{p}))^2}{\sigma^2(D_{ij}) + \sigma^2(M_{ij}) + \sigma^2(\text{sys})} \,. \tag{5.3}$$

$\sigma(\text{sys})$ is an estimated systematic error, set equal to a constant fraction, F_{sys}, of M_{ij}. This fraction, which is 5% in the tuning of Jetset 5.2 and 2% in the tuning

of Jetset 7.2, is set so that the best tune's χ^2/bin is approximately 1.0. $F_{\rm sys}$ is smaller for Jetset 7.2 because parton shower Monte Carlos such as Jetset 7.2 are better at reproducing experimental data than fixed order QCD Monte Carlos such as Jetset 5.2.

We can calculate this χ^2 for a variety of Monte Carlo parameter settings. For instance, we can choose to vary only one parameter at a time in order to obtain a crude optimization of the Lund model parameter set.

In this measurement, we tune in 5 parameters, and using this method to search for the minimum of χ^2 in this 5-dimensional parameter space, to several significant figures, would be time-consuming. Therefore, in order to obtain a fine tuning of the Lund model parameter set, along with an estimate of the associated systematic errors, a second method was used for comparing experiment to Monte Carlo. Instead of comparing experimental distributions to Jetset/TPCLUND distributions, this method unfolds the effect of detector acceptance on the experimental distributions and "fits" Jetset distributions (no TPCLUND) to these unfolded distributions, with bin entries D_{ij}.

The unfolding is accomplished by using a preliminary Lund parameter set, which we label $\vec{p_0}$, to generate a set of Jetset/TPCLUND distributions and a set of Jetset-only distributions. The ratio of the two distributions is used to unfold the experimental distributions.

The "fitting" is done as follows. A second set of Jetset distributions is generated using the preliminary parameter set $\vec{p_0}$ (the so-called *expansion point*). In addition, one parameter at a time is varied to each side of that parameter's preliminary value, and a set of distributions is generated at each of these points. Using these distributions, we can approximate the number of entries M_{ij} in each bin in each distribution as a first order Taylor expansion in the parameter space $\vec{p}$:

$$M_{ij}(\vec{p}) = M_{ij}(\vec{p_0}) + \vec{\nabla}_{\vec{p}} M_{ij}(\vec{p}) \mid_{\vec{p}=\vec{p_0}} \cdot (\vec{p} - \vec{p_0}) \tag{5.4}$$

The spacing of the parameter sets around the expansion point must be fixed at values that are not too small. Otherwise, the difference in the contents of some bins becomes smaller than the error in this difference, creating large random variations in the coefficient of $\vec{p} - \vec{p_0}$ in the second term of Equation 5.4 that tend to exaggerate greatly the growth of $M_{ij}(\vec{p})$ as $|\vec{p} - \vec{p_0}|$ increases. If this happens in many bins, then the minimum of χ^2 is forced to be artificially close to $\vec{p_0}$, with unreasonably small errors on this minimum.

We substitute Equation 5.4 into Equation 5.3 and solve for the parameter set that optimizes χ^2. Instead of summing over all 11 distributions, the old method summed over three distributions, one distribution chosen from each of the three sets of distributions in Table 5.4. Then χ^2 was computed 40 times, for the $40 = 5 \times 2 \times 4$

possible combinations of three distributions. The total χ^2 per bin is a measure of the goodness of the fit, the average of the 40 parameter sets is the new parameter set selected by the fit, and the σ of the 40 parameter sets is the systematic error of the parameter set.

In general, this new parameter set is not equal to the expansion point's parameter set, so this fitting process must be iterated, substituting this new parameter set for the old parameter set. This process is repeated until the change in the parameter set from one iteration to the next is less than the systematic error on the parameter sets.

5.3.3.3 Modifications to the Old Method.

A difficulty with the old fine tuning method is that convergence in the average of 40 parameter sets does not imply convergence in the individual parameter sets. The method used here combines all of the histograms in one of the 3 sets of histograms into a *superhistogram*, for all 3 sets, and the sum in Equation 5.3 is done over the 3 superhistograms. The tuning is then iterated until it is clear that the parameters are fixed, apart from statistical fluctuations whose size is estimated in the tuning. We consider the spread in the 40 parameter sets, as evaluated in the old method, a good estimate of the systematic error.

The old method ignored $\sigma^2(M_{ij})$ in Equation 5.3, thus making it possible to solve for the minimum in χ^2 using a system of linear equations. The method used here does not ignore this contribution to the error. Taking this contribution into account sometimes causes the minimum of χ^2 to correspond to physically absurd parameter sets, due to statistical errors in Equation 5.4. This pathological behavior is cured by using the common trick of finding the minimum of χ^2 in a stepwise manner. In each step, χ^2 is minimized with $\sigma^2(M_{ij})$ fixed at the value calculated using the $\vec{p}$ determined in the previous step. For the first step, the expansion point $\vec{p}_0$ is used to calculate $\sigma^2(M_{ij})$. This process is repeated until $\vec{p}$ converges, typically after 4-6 steps.

This method should give the same value of $\vec{p}$ as the coarse tuning method discussed above if enough time were spent on the latter method. We do not attempt to do a fine tune using the coarse tuning method.

5.3.3.4 Backgrounds.

There is an additional pitfall in the tuning: backgrounds. The τ and 2γ backgrounds are estimated to be approximately 1% of the good hadronic event sample, so one might naively think that backgrounds can be ignored. They can not be ignored because they are concentrated in small regions of some of the tuning distributions where there are not many hadronic events. τ and 2γ events have low

multiplicities, so those τ and 2γ events that pass the hadronic event selection cluster in the lowest bins in multiplicity. τ events, because of the low τ mass, are highly collimated and back-to-back, so they also cluster at large thrust, low thrust major, and low thrust minor. During the tuning procedure, before the τ-2γ selection was implemented to eliminate these backgrounds (described in Section 5.2), data clearly showed a statistically significant excess over Monte Carlo for multiplicities of 5 and 6, and especially for the lowest bins in thrust minor where there are essentially no hadronic events. A less significant excess was seen at low thrust major.

We did not try to simulate these backgrounds for two reasons. First, the simulation of 2γ events is difficult. Second, according to TPCLUND6, approximately half of the τ events that passed the Good Hadronic Event Selection have nuclear interactions, and the TPCLUND nuclear interaction simulation is not considered very reliable.

The alternative to simulating these backgrounds is to cut them. To avoid losing information for the fitting, we did not remove from the fit those bins where the backgrounds are important. Instead, we decided to eliminate the backgrounds by designing selection criteria in variables designed especially to remove these backgrounds. The result is the τ-2γ selection of Section 5.2.

TPCLUND6 estimates that the τ-2γ selection removes 94% of all τ events that pass the good hadronic selection, 37% of the passing hadronic events with 8 or less good hadrons, and 11% of all the passing hadronic events.

The τ-2γ selection removes or makes statistically insignificant the previously mentioned differences. Figure 5.1 shows the differences, in standard deviations, between tuned Monte Carlo and Experiment 14-18 data before and after the τ-2γ selection was implemented. The tuned parameters did not change a large amount after the τ-2γ selection was implemented, which is not surprising, since few bins were affected. However, the χ^2/bin, with the systematic error floor $F_{\text{sys}} = 2\%$, dropped from about 1.2 to about 1.0.

5.3.3.5 Results of the Tuning.

For the tuning of Jetset 7.2, b was fixed by tuning it so that Jetset reproduced the TPC D^* momentum spectrum. The D^* spectrum is now controlled by ϵ_c, so we can and do tune both a and b. In fact, we have simultaneously tuned Λ_{QCD}, Q_0, a, b, and σ_q with two different values of the systematic error floor F_{sys}: 2% and 0%. The optimizations were well-behaved. The results are in Table 5.5, and histograms of the number of events or tracks as a function of the the tuning variables for experimental data and for Monte Carlo with $F_{\text{sys}} = 2\%$ are in Figures 5.2 through 5.4. The tune with $F_{\text{sys}} = 2\%$ is our best tune of the Monte Carlo. We use the tune with $F_{\text{sys}} = 0\%$ for estimating systematic errors.

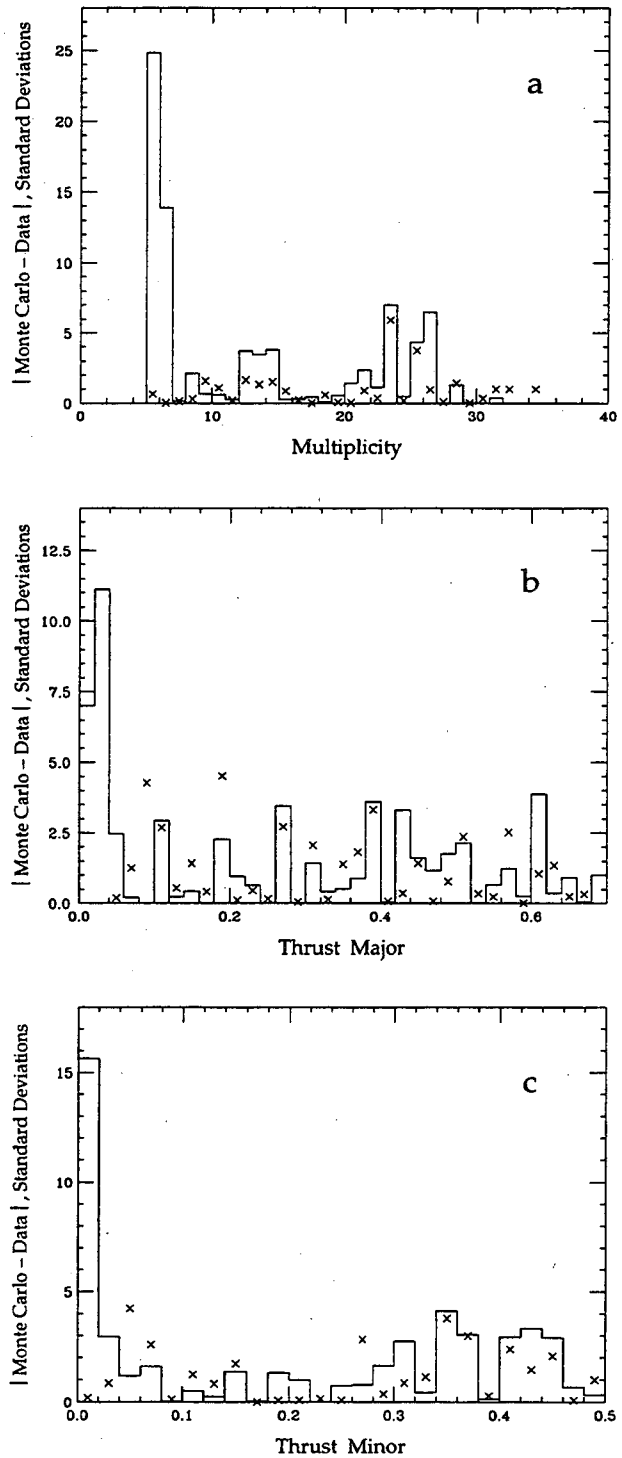


Figure 5.1: Absolute value of the discrepancy, in standard deviations, between Monte Carlo and Experiment 14-18 data without (solid) and with (crosses) the τ-2γ selection for multiplicity N (a), thrust minor L_1 (b), and thrust major L_2 (c).

F_{sys}	Parameter					χ^2/bin
	Λ_{QCD}	Q_0	a	b	σ_q	
2%	.364 ± .022	1.360 ± .129	.053 ± .051	.497 ± .071	.320 ± .006	0.94
0%	.317 ± .016	1.870 ± .346	.385 ± .049	.632 ± .055	.336 ± .005	1.33

Table 5.5: Tune of the Lund parameters in Jetset 7.2.

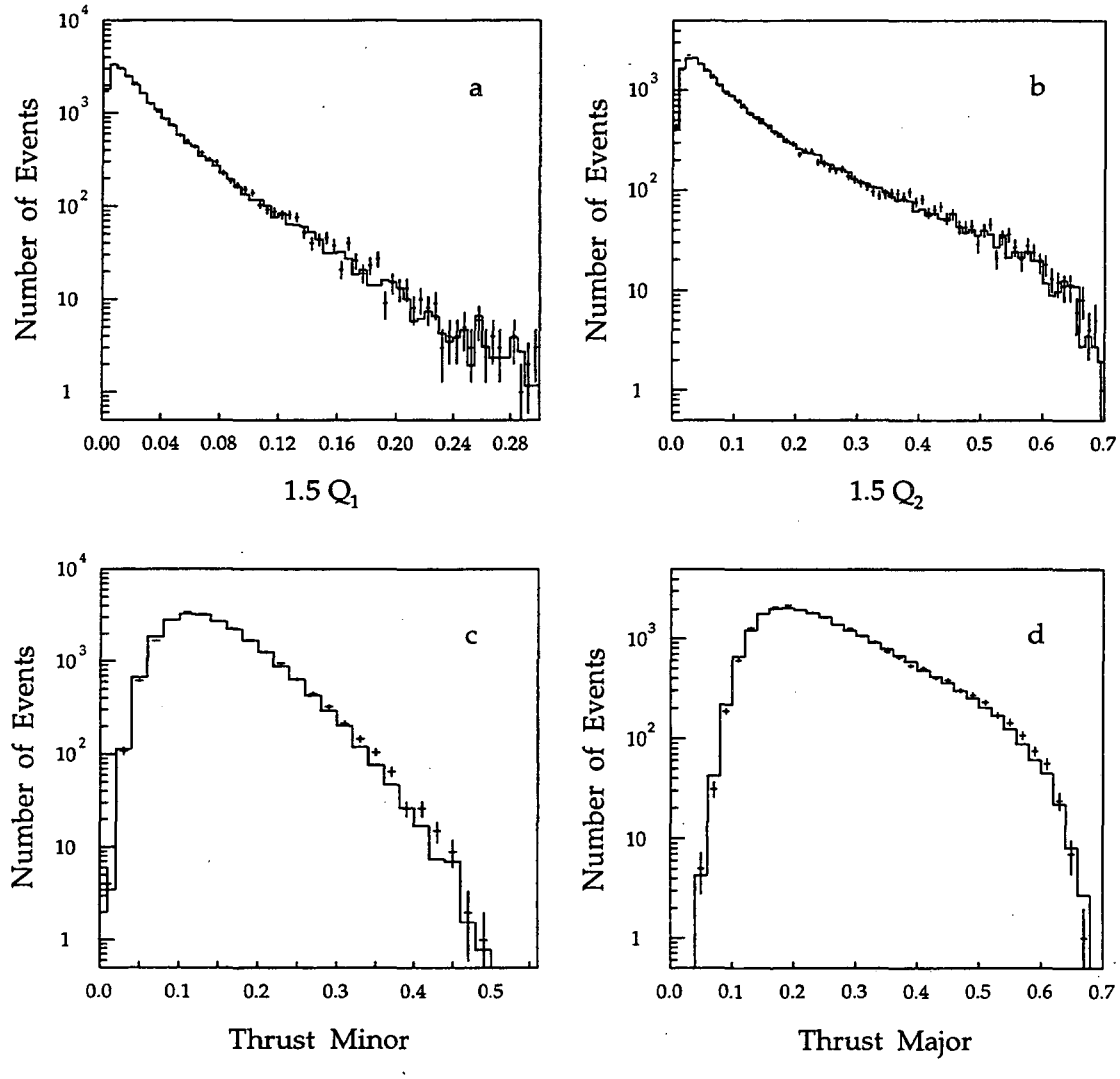


Figure 5.2: Experimental data (points) and tuned Monte Carlo (solid) histograms of the number of events as a function of the aplanarity $1.5Q_1$ (a), $1.5Q_2$ (b), the thrust minor L_1 (c), and the thrust major L_2 (d).

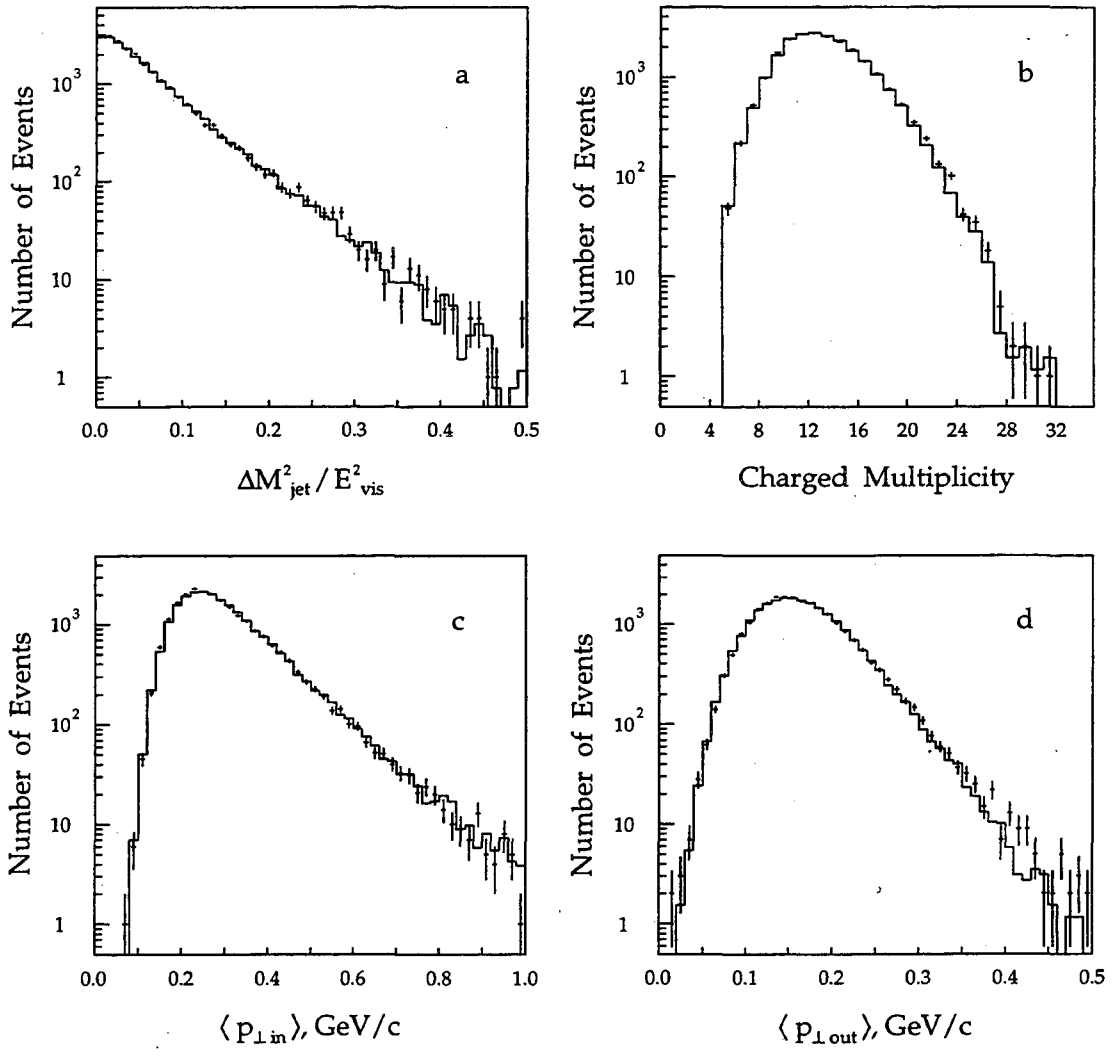


Figure 5.3: Experimental data (points) and tuned Monte Carlo (solid) histograms of the number of events as a function of $\Delta M^2_{jet}/E^2_{vis}$ (a), Charged Multiplicity (b), $\langle p_{\perp in} \rangle$ (c), and $\langle p_{\perp out} \rangle$ (d).

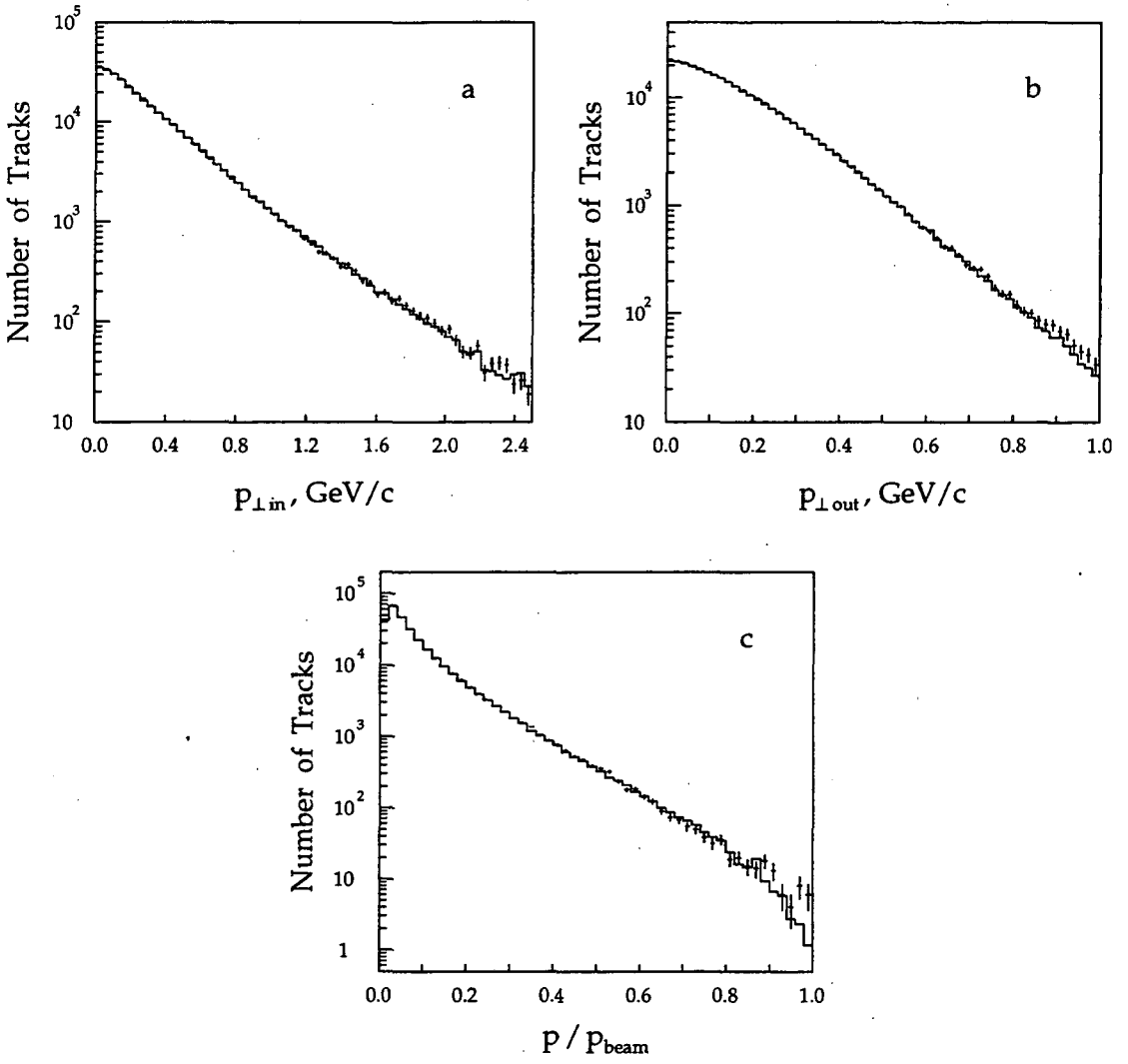


Figure 5.4: Experimental data (points) and tuned Monte Carlo (solid) histograms of the number of tracks as a function of $(p_i)_{\perp in}$ (a), $(p_i)_{\perp out}$ (b), and x_p (c).

The big difference between the 2 sets of parameters is that Lund a and b are higher for the tuning with the 0% systematic error floor than for the tuning with the 2% systematic error floor. The reason this happens is that the track momentum spectrum for $x_p > .14$ is best fit by small values of a and b, while $x_p < .14$ is best fit by larger values of a and b.

The momentum is binned in equal intervals in x_p, thus most of the bins have $x_p > .14$, but these bins contain only 19% of the tracks, as shown in Figure 5.4c. The higher statistics for bins with $x_p < .14$ favor these bins, while the greater number of bins for $x_p > .14$ favors these bins. Tuning with a 2% systematic error floor dilutes the weighting of the high statistics for $x_p < .14$, so the 2% tune fits $x_p > .14$ better than $x_p < .14$, resulting in small a and b. Tuning with a 0% systematic error floor does not dilute $x_p < .14$, resulting in larger a and b.

Chapter 6

Feed-Forward Neural Networks

In this chapter, we describe the design, training, and use of feed-forward neural networks[1]. We also describe some relevant prior applications of feed-forward neural networks to high energy physics. In Chapter 7, we describe our use of a neural network to measure $\sigma(b\bar{b})/\sigma(q\bar{q})$, and in Chapter 8 we describe our use of a neural network in a study of bias in measurements of charged hadron production in bottom jets, where neural networks are used to tag the bottom-quark jets.

6.1 Neural Network Architecture

A feed-forward neural network is pictured as a layered array of nodes with a connection between each pair of nodes in adjacent layers, as shown in Figure 6.1. We shall use as a general example a network with three layers: an input layer (the bottom row of nodes in Figure 6.1), an output layer (the top row), and one *hidden* layer (the middle row). A network can have any number of hidden layers, including zero. It has been shown that there is nothing to be gained from having more than two hidden layers, and that one hidden layer is often sufficient [94, 95].

Each node represents a number: x_k for node k in the input layer, h_j for node j in the hidden layer, and y_i for node i in the output layer. For a general type of analysis, where one wishes to classify an N-dimensional vector as belonging to one of L classes based upon differences in how the classes are distributed in the N-dimensional space of all vectors, one has N nodes in the input layer, each node representing one of the components of the vector. Each input N-tuple is called a *pattern*.

[1]Reference [93] is a recent review of neural networks, accessible to readers without previous knowledge of neural networks, along with an extensive bibliography and a guide to the neural network literature.

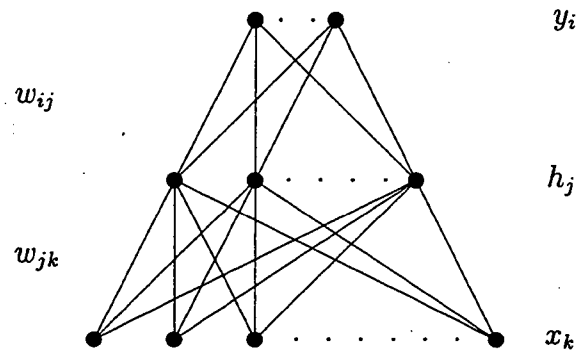


Figure 6.1: A one hidden layer feed-forward neural network architecture (From [96]).

The M output nodes are used to classify the input pattern. For a one output network, the value of the output is a simple yes (1) or no (0) if the classification of the input pattern is certain, and in between 0 and 1 if the classification is uncertain. For more than one output node, each output node's value can encode one bit of information in an M-bit binary word representing the network's classification of the input pattern, and the network can make 2^M classifications. Alternatively, each node can represent one of M classifications: the output node representing the assigned classification is 1 and all of the other nodes are 0. In any case, each classification of the input pattern has a corresponding target output M-tuple $(y_1, ..., y_M) = (t_1, ..., t_M)$.

We now describe how the values of the numbers represented by each node are set. The values of the numbers in the input nodes are set equal to the values of the corresponding components of the pattern being presented to the network. In the other layers, the number in each node in a particular layer is a function of the numbers in all of the nodes in the previous layer. For the output layer, this function is

$$y_i = g\left[a_i\right] = g\left[\theta_i + \frac{1}{T}\sum_j \omega_{ij} h_j\right] , \tag{6.1}$$

while for the hidden layer, the function is

$$h_j = g\left[a_j\right] = g\left[\theta_j + \frac{1}{T}\sum_k \omega_{jk} x_k\right] , \tag{6.2}$$

where T is the network *temperature* (an inverse gain), ω_{ij} (ω_{jk}) is the weight (or strength) of the connection between nodes i (j) and j (k), θ_i (θ_j) is the threshold for

node i (j), and $g(a)$ is a non-linear *transfer function.* Typically, transfer functions asymptotically approach a fixed value for argument $a \rightarrow \infty$ and another fixed value for $a \rightarrow -\infty$, a common choice is $g(a) = \tanh(a)$. In this fashion, numbers are propagated forward through the network from the input layer to the output layer, hence the name of this type of network: *feed-forward.*

In Equations 6.1 and 6.2, the argument of g is a linear function of the values of the nodes in the previous layer. When this argument, $\theta_j + \sum_k \omega_{jk} x_k$, is set equal to a constant, it defines a hyperplane in the space of vectors defined by $\vec{x} = (x_1, ..., x_k, ...)$. Therefore, a 0-hidden-layer neural network maps regions of the input space bounded by hyperplanes to small regions of the space of outputs, and such networks are functionally identical to a linear discriminant. The non-linearity of g enables networks with at least one hidden layer to map regions of the input space with curved boundaries to small regions of the output space. Thus, neural networks are often more powerful than linear discriminants or other multi-dimensional classifiers [97].

6.2 The Training of a Neural Network

A neural network does not automatically classify the input patterns. The parameters (weights and thresholds) of the network must be set to values that enable the network to perform this task. The process in which these parameters are set is called *training.*

The most common method for training uses the *back-propagation* algorithm [98], which is used in this analysis. Training a network requires a *training set* of L equal-sized samples of patterns, one sample from each classification. Equal-sized samples are used so that the network does not favor one class over the others.

In the original *off-line* use of the back-propagation algorithm, the entire training set is presented to the network repeatedly, each complete presentation being called an *epoch.* At the end of each epoch, the weights and thresholds are incremented to reduce the error

$$E = \frac{1}{2} \sum_p \sum_i (y_i^{(p)} - t_i^{(p)})^2 \ , \tag{6.3}$$

where $y_i^{(p)}$ and $t_i^{(p)}$ are the calculated and target output values in node i for pattern p.

Suppose that there are n weights. The current value of these weights can be organized into an ordered n-tuple that is a point P in the real r-dimensional Euclidean space $\mathbf{R}^n$ of all possible n-tuples. Then the rule for incrementing the weights is for the point P to move opposite the direction of the gradient of E in

the space of weights:

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial \omega_{ij}} \tag{6.4}$$

and

$$\Delta\omega_{jk} = -\eta \frac{\partial E}{\partial \omega_{jk}} \ , \tag{6.5}$$

where η is the *learning rate*. The partial derivatives are easily calculated using the chain rule:

$$\frac{\partial E}{\partial \omega_{ij}} = (y_i - t_i) g'(a_i) h_j / T \tag{6.6}$$

and

$$\frac{\partial E}{\partial \omega_{jk}} = \sum_i (y_i - t_i) g'(a_i) g'(a_j) \omega_{ij} x_k / T^2 \ . \tag{6.7}$$

Similar formulæ exist for the thresholds. The initial values of the weights and thresholds must be randomly chosen; otherwise, the training may have difficulty breaking the symmetry of the initialization (e.g. all parameters zero). Typically, these initial values are randomly generated uniformly in an interval $[-\omega_0, \omega_0]$.

The weights and thresholds are repeatedly incremented in this way in a stepwise search for the minimum of E in the space of weights and thresholds, until E no longer decreases. There are ways of avoiding local minima [96]. In minimizing E, the actual network output M-tuple $(y_1, ..., y_M)$ for each class of patterns is forced to be as close as possible to that class's target output M-tuple, and the different classes of patterns are separated from each other as much as possible. This is how the network gains its ability to distinguish the classes of inputs from each other.

In practice, *on-line* training is used. Instead of updating the weights and thresholds at the end of each epoch, small groups of patterns composed of equal-sized sets of randomly selected events from each class of patterns are presented to the network, and the network parameters updated after each group has been input. In addition, the updating formulæ are changed to

$$\Delta\omega_{ij}(t) = -\eta \frac{\partial E}{\partial \omega_{ij}} + \alpha \omega_{ij}(t-1) \tag{6.8}$$

and

$$\Delta\omega_{jk}(t) = -\eta \frac{\partial E}{\partial \omega_{jk}} + \alpha \omega_{jk}(t-1) \tag{6.9}$$

to damp out oscillations. The new second term in these equations is a momentum term proportional to the increment of ω in the previous update; t is the update number and α should be between 0 and 1. On-line back-propagation better avoids local minima and is often faster than off-line back-propagation.

The decision about when to terminate training requires a *test set*, which is like the training set in that it also consists of L equal-sized samples of patterns, one

sample from each classification, but the patterns in the training and tests sets are different. What is important is not the network's ability to classify the training set patterns, but rather the network's ability to generalize what it learned from the training set to the task of classifying an independent test set's patterns.

It often happens that the error E evaluated for a test set will decrease to a minimum and then actually starts to increase. This happens because, at first, the network is learning the general characteristics that distinguish the patterns in each classification from each other, but past a certain point the network learns the statistically random peculiarities that distinguish the training set's classifications from each other. This is called *overtraining*. Overtraining is handled by increasing the size of the training set, thereby reducing statistical fluctuations and increasing the network's performance on the test set. Increasing the size of the training set also decreases the performance on the training set towards the performance on the test set, since the network can not learn the training set's peculiarities as readily. If the size of the training set becomes too small, then the network memorizes individual patterns and the network's ability to generalize plunges.

On the other hand, if the ratio of the number of training set patterns to the number of network parameters is too big, the probability of getting stuck in a local minimum during training increases. A good compromise between these competing demands is for this ratio to be on the order of 100 [99].

The issue of the choice of the number of nodes in each hidden layer can now be handled in several ways. There exist a variety of algorithms designed to be used to eliminate unnecessary nodes. A class of such algorithms are *pruning* algorithms, [96] which simultaneously train the network and set to zero the weights connected to unneeded nodes. Our experience shows that the number of nodes the algorithm decides to prune is not directly enough related to the pruned network's performance. Instead, we take a direct approach: we see how the performance of the network varies as a function of the number of hidden nodes. Typically, the performance is seen to rise rapidly with increasing number of hidden nodes when this number is small (see Section 7.2). The performance then asymptotically approaches a limiting value as the number of nodes becomes larger. Measures of network performance besides E are discussed in the next section.

The traditional approach in High Energy Physics of doing analyses with a sequence of selections is equivalent to a structured decision tree, where decisions are made sequentially. In contrast, a neural network is a form of parallel computing. Each node in a network "votes" on the network's decision. As a result, neural networks have two desirable properties that decision trees do not have [99]:

1. Robustness: a few bad inputs can be tolerated in decision-making.

2. Generalizability: networks respond to previously unseen patterns by generalizing from what the network learned during the training process.

These properties, and the power and general applicability of neural networks, are the reason for the interest in neural networks in the field of High Energy Physics. On the other hand, the voting of the nodes means that individual nodes do not have specific functions. A network's decision-making is spread among all of the nodes.

6.3 Measuring Network Performance

What we mean by a network's performance is its ability to distinguish patterns from different classes. The usual measure of performance is the error E, but there exist other measures of performance. For example, with a one output network, it is common to calculate the efficiency and purity, as a function of a cut on the network output, for the signal patterns in the sample of patterns with outputs above the cut. We have chosen to use a compact measure of performance for a one output network that directly quantifies the statistical separation of the classes [100, 101].

This method, for the case of two classes ($L = 2$ in the notation of Section 6.1), uses a set of signal patterns, a set of background patterns, and a data set composed of signal and background only. As an example, the signal is composed of bottom events, the background is composed of non-bottom events, and data is composed of a mixture of the two. We start out by producing neural network output distributions for these three sets of patterns. Then, a fit is done of a linear combination of the signal distribution and the background distribution to the data distribution, yielding the error on the fraction of the data that is fitted to be signal. Our measure of network performance is the dependence of this error on the overlap of the two distributions, with the dependence of the error on the finite size of the signal and background samples removed.

We work out what our measure of performance is mathematically. Binning the network output and labelling the bins i, let s_i, b_i, and d_i be the fraction of the signal, background, and data (respectively) in bin i, let N be the number of data patterns, and let α be the fraction of the data that is actually composed of signal. Then the χ^2 of the fit of a linear combination of signal and the background to the data is

$$\begin{aligned} \chi^2 &= N \sum_i \frac{(\alpha s_i + (1-\alpha) b_i - d_i)^2}{\sigma^2(d_i)} \\ &= N \sum_i \frac{(\alpha s_i + (1-\alpha) b_i - d_i)^2}{d_i} \, . \end{aligned} \tag{6.10}$$

The value of α that minimizes χ^2 is

$$\hat{\alpha} = \frac{\sum_i \frac{(d_i - b_i)(s_i - b_i)}{d_i}}{\sum_i \frac{(s_i - b_i)^2}{d_i}} , \tag{6.11}$$

with

$$\sigma(\hat{\alpha}) = \sqrt{\frac{1}{N \sum_i \frac{(s_i - b_i)^2}{d_i}}} \tag{6.12}$$

$$= \sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{NF}} , \tag{6.13}$$

where we have defined

$$F = \hat{\alpha}(1-\hat{\alpha}) \sum_i \frac{(s_i - b_i)^2}{d_i} . \tag{6.14}$$

In Monte Carlo studies, we make the substitutions $\hat{\alpha} = \alpha$ and $d_i = \hat{\alpha} s_i + (1-\hat{\alpha}) b_i$.

F is our measure of network performance. If we could achieve perfect separation of signal and background (that is, each bin i has either $s_i = 0$ or $b_i = 0$), then $F = 1$. If $s_i = b_i$ for all i, then $F = 0$. Otherwise, $0 < F < 1$.

The generalization to more than two classes of patterns ($L > 2$) is derived in Appendix B.

An example of an application of F is in Reference [101], which compares the performance of networks with different types of inputs. One approach is to use, as inputs, 'raw' quantities such as particle 4-momenta, 3-momenta, or energy deposits in a calorimeter [99, 101, 102, 103]. The motivation for using 'raw' quantities is to provide the network with as much information as possible and to have the network extract the useful information.

Another approach is to use, as inputs, constructed *shape variables* that describe the event shape (e.g. thrust, sphericity, and thrust minor) that have different distributions for b and non-b events. The preprocessing of the information into these shape variables ought to make it easier for the network to extract the information useful for distinguishing b and non-b events. There are a large number of possible shape variables that can be used as inputs, so we can pick a small set of variables such that the network output distributions for b and non-b events are as well separated as possible [99, 101, 104, 105, 106, 107].

The raw input approach keeps information that might otherwise be thrown away in the preprocessing approach. As a result, one might expect that networks with raw inputs perform better, but with the trade-off that raw input networks take much longer to train. For the examples studied in Reference [101], these expectations are fulfilled. In this reference, the preprocessed 25-input network has $F = 0.258$ and the raw 23-input network has $F = 0.312$; both networks have one hidden layer containing 5 nodes.

6.4 Previous Uses of Neural Networks in High Energy Physics

Neural networks are increasingly being used in high energy physics [108, 109, 110]. Previous applications of back-propagation neural networks to high energy physics can be roughly divided into two groups: data compression for fitting in a network output, and classification.

6.4.1 Classification Using Neural Networks

In an example of this application, calorimeter information is used for inputs to a neural network, which attempts to determine what kind of particle generated the calorimeter information [111, 112, 113]. There is one output for all of these networks. A cut is made on this output and all patterns that produce an output above this cut are classified as being produced by one kind of particle, while all other patterns are classified as being produced by another particle. Figure 6.2 shows the network output distributions for electrons and for hadrons for a network designed to distinguish these two classifications of particles [111]. These two distributions are well separated; for this sort of application the two output distributions must have little or no overlap.

6.4.2 Fitting in a Neural Network Output

Analyses in which the $b\bar{b}$ width of the Z^0 is measured are the archetype of this approach. Here, a linear combination of the neural network output distributions for Monte Carlo b events and for Monte Carlo non-b events is fitted to the network output distribution for experimental data [114, 115]. Figure 6.3 shows the L3 Collaboration's fit and the distributions for b events and for non-b events. Both Monte Carlo distributions fill the entire range of network outputs, from 0 to 1, which is what happens when the network inputs do not carry enough information to distinguish unambiguously the two types of events. It may even happen that the output distributions do not cover the entire range from 0 to 1. This is the method used in this analysis to measure the fraction of hadronic events that are bottom events.

An inherent disadvantage of this type of analysis is the reliance upon Monte Carlo for training the networks. This reliance is necessary because there are no unbiased samples of b or non-b events available on which to train a network.

It is also possible to do this kind of analysis with three or more classes of patterns. The DELPHI Collaboration [116] has measured the bottom, charm, and light (up, down, and strange) quark widths of the Z^0 using a 3-output network.

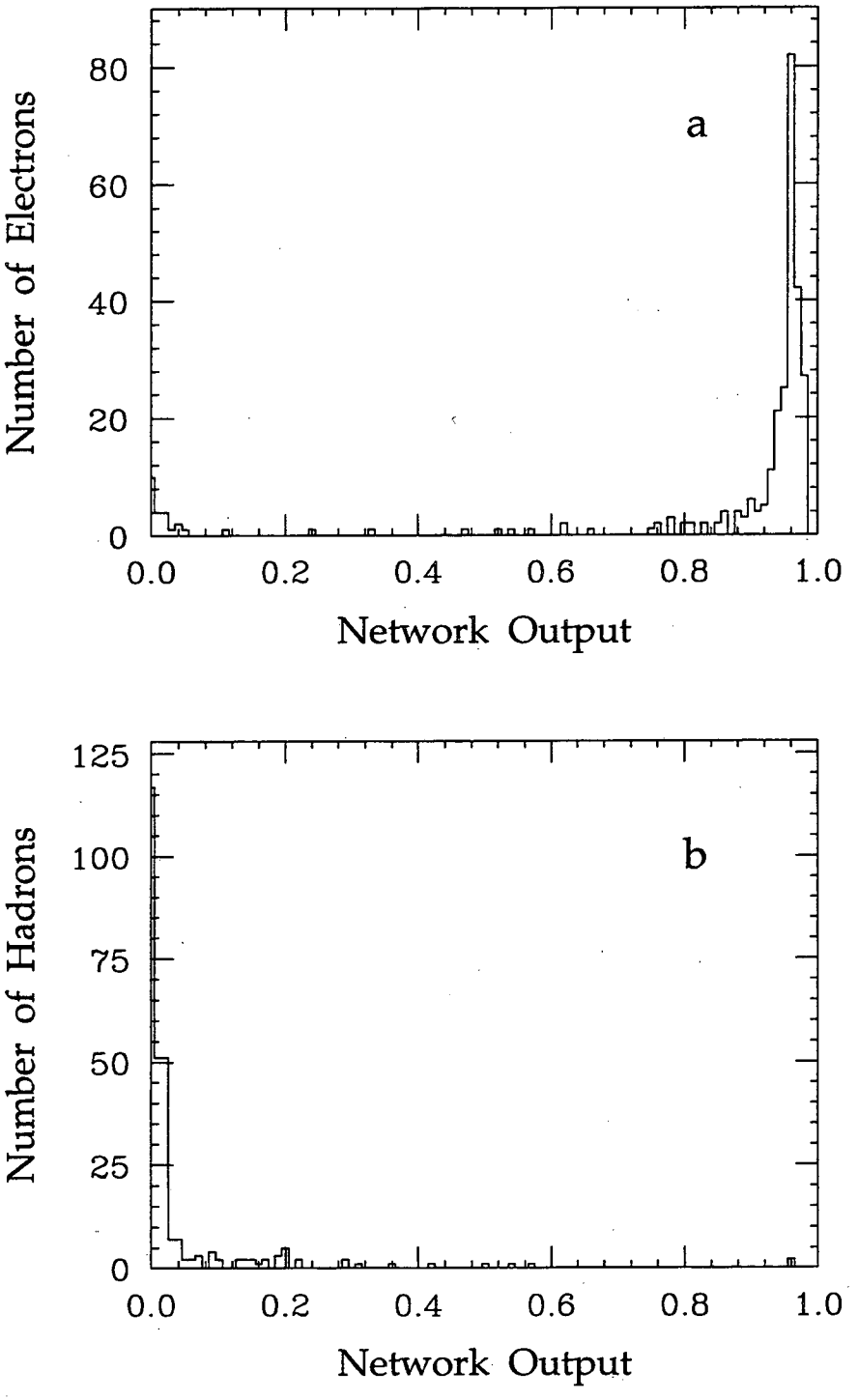


Figure 6.2: Neural network output distributions for electrons (a) and for hadrons (b) for the DØ neural network designed to distinguish the two (From [111]).

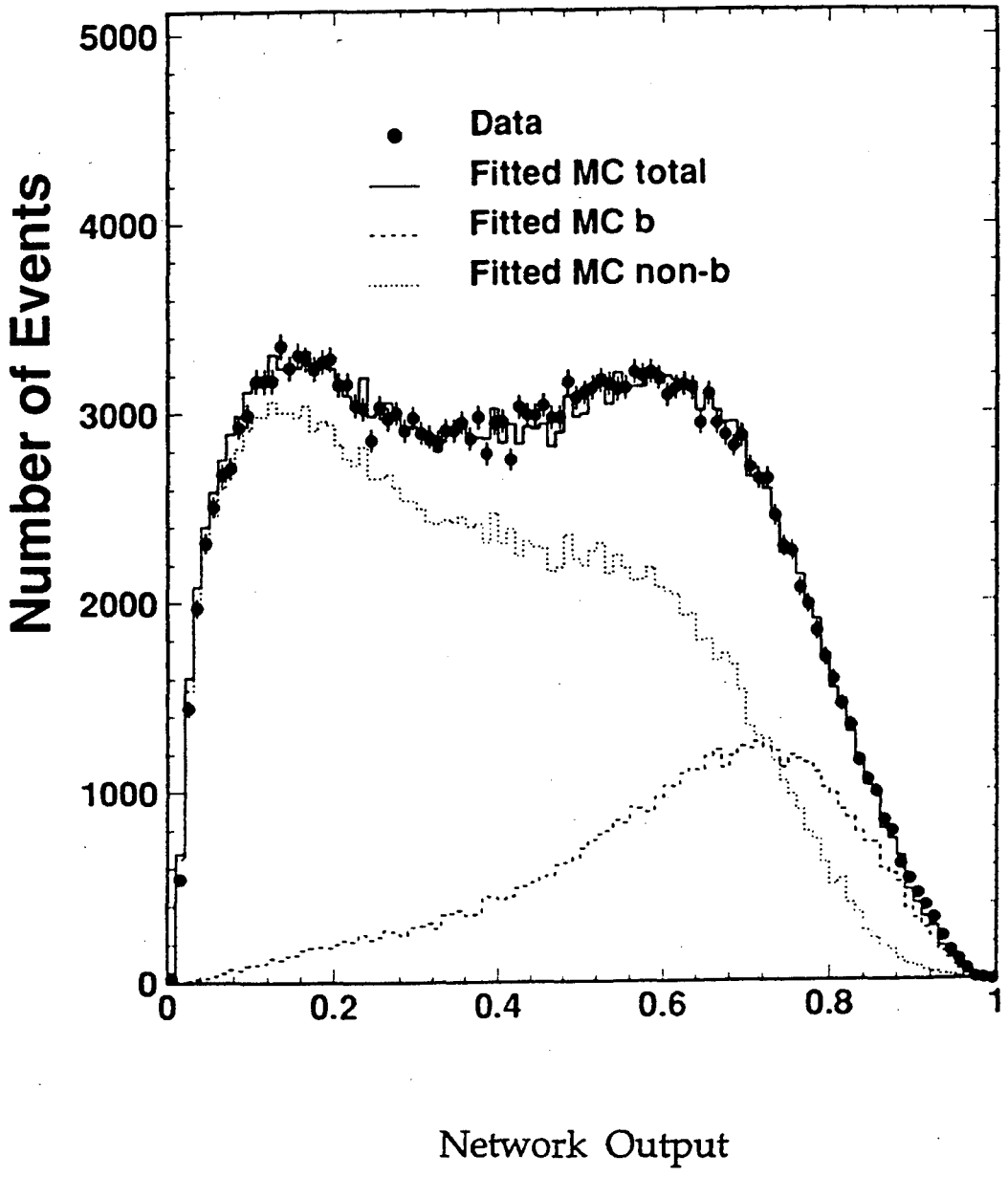


Figure 6.3: L3 fit of a linear combination of the neural network output distributions for Monte Carlo b and Monte Carlo non-b events to the network output distribution for experimental data (From [115]).

Dalitz-type scatterplots of the output distributions are shown in Figure 6.4, with the sum of the three outputs normalized to 1. As discussed in Section 6.1, only two independent outputs are needed for this analysis. It is possible to use a 1-output network and fit a linear combination of three distributions, but it would not be possible to train the network to recognize the three classes of events. Instead the network would be trained on sets of b and non-b events and it would respond differently to charm and light quark events.

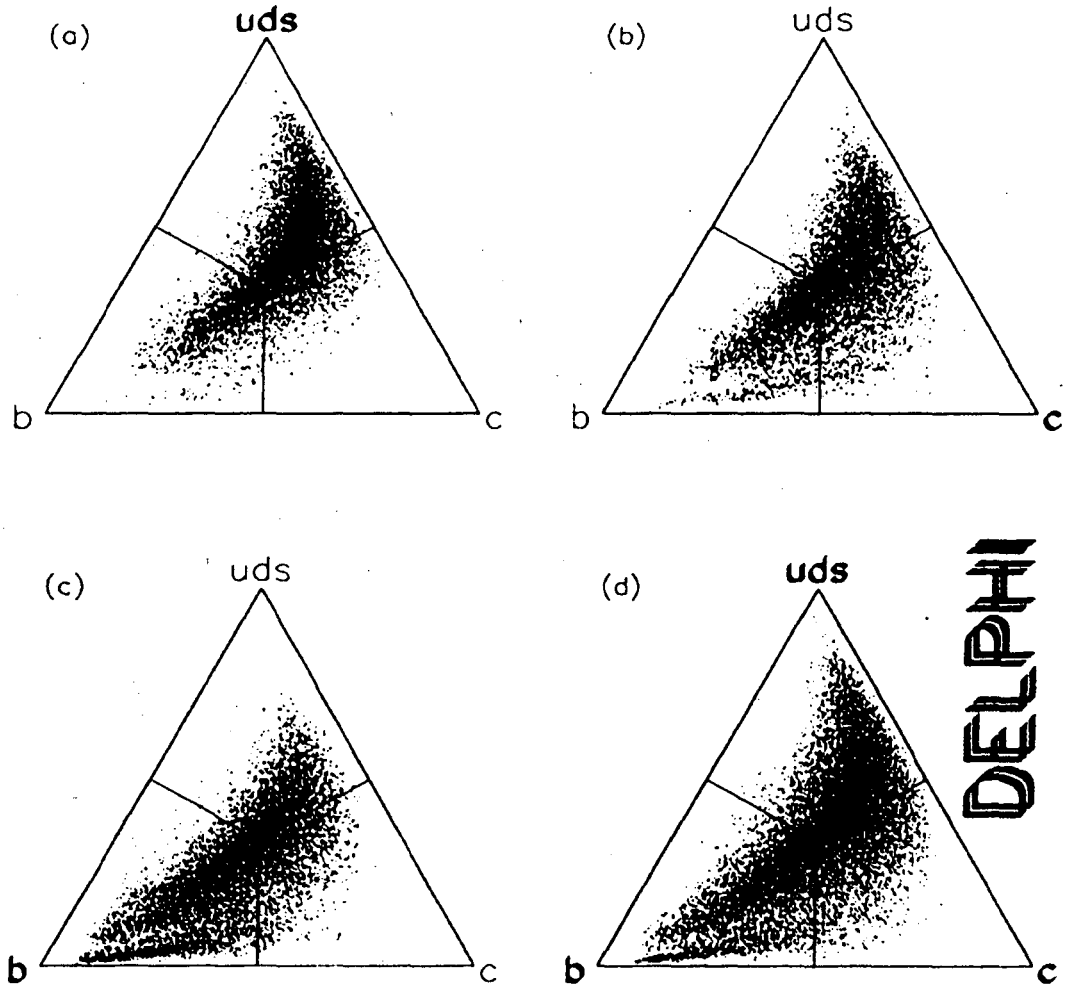


Figure 6.4: DELPHI neural network output distributions for Monte Carlo uds events (a), c events (b), and b events (c), and for experimental data (d) (From [116]).

Chapter 7

A Measurement of the Bottom Event Production Fraction

The ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$ has never been measured in the continuum. Normally, the value of $\sigma(b\bar{b})/\sigma(q\bar{q})$ is deduced from the discontinuity in $R = \sigma(q\bar{q})/\sigma(\mu^+\mu^-)$ at the $b\bar{b}$ threshold. In this chapter, we present a first measurement of the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$ in the continuum. This measurement is based on 66 pb^{-1} of TPC/2γ data collected between 1984 and 1986, and it uses a neural network with inputs that are computed from charged hadron 3-momenta to distinguish b and non-b events.

In this chapter, we first describe the network's inputs and architecture, and how they were designed. Next, we show how the network was trained. We then describe and implement a method for fitting the bottom event fraction in the experimental data. We describe how the fitted bottom event fraction depends upon the Monte Carlo bottom event fraction, and we use a new method to extract a consistent bottom event fraction. Finally, we discuss our measurement of $\sigma(b\bar{b})/\sigma(q\bar{q})$ and compare it to previous measurements of $\Gamma(Z^0 \rightarrow b\bar{b})$.

7.1 The Choice of Neural Network Inputs and Architecture

For the *event-tagging neural network* we constructed to identify bottom events, we chose the preprocessed input approach discussed in Section 6.3, for the sake of simplicity. We compiled a list of candidate inputs and for each of them computed F (described in Section 6.3), which quantifies how well the b and non-b distributions are separated. We selected 7 variables[1], uses a with relatively large values of F

[1] $N = 7$ in the notation of Section 6.1

that are fairly independent of each other. The selected variables, computed using the tracks and events that pass the selections described in Section 5.2, are:

1. $N\,E_{cm}/E_{vis}$, the scaled event charged multiplicity, where N is the number of charged tracks in the event and E_{vis} is the event visible energy.

2. $\sum p_{\perp}^{out}\,E_{cm}/E_{vis}$, the scaled sum, over all tracks, of the component of the track momenta perpendicular to the event plane, the plane defined by the thrust axis and the thrust major axis (these axes are defined in Section 5.3.3.1).

3. $\sum p_{\perp}^{in}\,E_{cm}/E_{vis}$, the scaled sum, over all tracks, of the component of the track momenta in the event plane and perpendicular to the event (thrust) axis.

4. $\sum p_{\|}\,E_{cm}/E_{vis}$, the scaled sum, over all tracks, of the component of the track momenta parallel to the event axis.

5. $M^{(1)}\,M^{(2)}\,E_{cm}^2/[4E_{vis}^{(1)}\,E_{vis}^{(2)}]$, the product of the scaled invariant masses of both hemispheres, where the hemispheres are divided by the plane perpendicular to thrust axis, $M^{(i)}$ is the invariant mass of hemisphere i, and $E_{vis}^{(i)}$ is the visible energy of hemisphere i.

6. $p_{\|1}^{(1)} \times p_{\|1}^{(2)}$, where $p_{\|1}^{(i)}$ is $p_{\|}$ for the leading track in hemisphere i.

7. The boosted sphericity product $S_1 \times S_2$, where S_i is the sphericity of hemisphere i calculated in the frame of reference boosted by $\beta = .47$ along the event axis, in the direction of hemisphere i. $\beta = .47$ was chosen because it maximizes F for this variable.

The distributions of these variables for b and non-b Monte Carlo events that pass the analysis selections are shown in Figures 7.1 and 7.2, and the values of F for the 7 inputs are listed in Table 7.1. Note that these variables take advantage of the differences between bottom and non-bottom events discussed in Section 2.5.

To help distinguish bottom events from 3-jet non-bottom events, we use as inputs $\sum p_{\perp}^{in}\,E_{cm}/E_{vis}$ and $\sum p_{\perp}^{out}\,E_{cm}/E_{vis}$, since the latter is considerably smaller than the former for 3-jet events, which have a distinctly flat shape, whereas bottom events are more isotropic in the plane perpendicular to the event axis.

The boosted sphericity product, $S_1 \times S_2$, also distinguishs bottom events from 3-jet non-bottom events. Bottom events have large $S_1 \times S_2$, since the tracks in each of the two hemispheres are largely produced by the roughly isotropic decay of a bottom hadron with a boost in the neighborhood of $\beta = .47$, giving S_1 and S_2 large values. In contrast, 3-jet non-bottom events often have one hemisphere

Input:	1	2	3	4	5	6	7
F:	0.080	0.100	0.155	0.120	0.130	0.058	0.129

Table 7.1: The values of F for the 7 event-tagging network inputs.

with large S, since that hemisphere contains 2 jets, while the other hemisphere has small S, since it contains one non-bottom jet.

The inputs we use are these 7 variables after they have been linearly scaled so that the range of values for each input is 0 to 1. If we do not scale the inputs, the neural network inputs can have significantly different ranges of values, and this can reduce the network's performance. This reduction happens because the network parameters are normally initialized, before training has started, so that they are randomly distributed in a fixed interval, and if the ranges of the inputs are very different, training will favor some of the inputs [103].

To distinguish two classes of events from each other, the event-tagging network has one output. The target output value is 1 for bottom events and 0 for non-bottom events. The network has one hidden layer with 4 nodes. How the number of nodes in the hidden layer was picked is discussed in Section 7.2.

7.2 Training the Event-Tagging Neural Network

As first discussed in Section 6.2, a network's weights and thresholds (parameters) must be determined through training in order to make the network output distinguish bottom and non-bottom events, and to do so most effectively.

We use Jetnet 2.0 [96] as the software implementation for our neural network. The following Jetnet 2.0 default settings were chosen: the learning rate η was set to 0.01, the momentum term coefficient α was set to 0.5, the temperature T was 1.0, the transfer function $g(a)$ was chosen to be $1/[1+\exp(-2a)]$, and the initial values of the weights and thresholds were randomly chosen in the interval $[-0.1, 0.1]$.

There are two additional choices for the network that were optimized: the number of nodes in the hidden layer, and the ratio of the number of patterns to the number of network parameters. To determine the number of hidden layer nodes, we trained a number of networks, each with a different number of hidden nodes (Figure 7.3); the training process is described below. The chosen number is 4, which maximizes F. We also trained a number of networks, each with a different ratio of the total number of training patterns to the number of network parameters (Figure 7.4) and found the best ratio to be 200. These two choices are independent of each other.

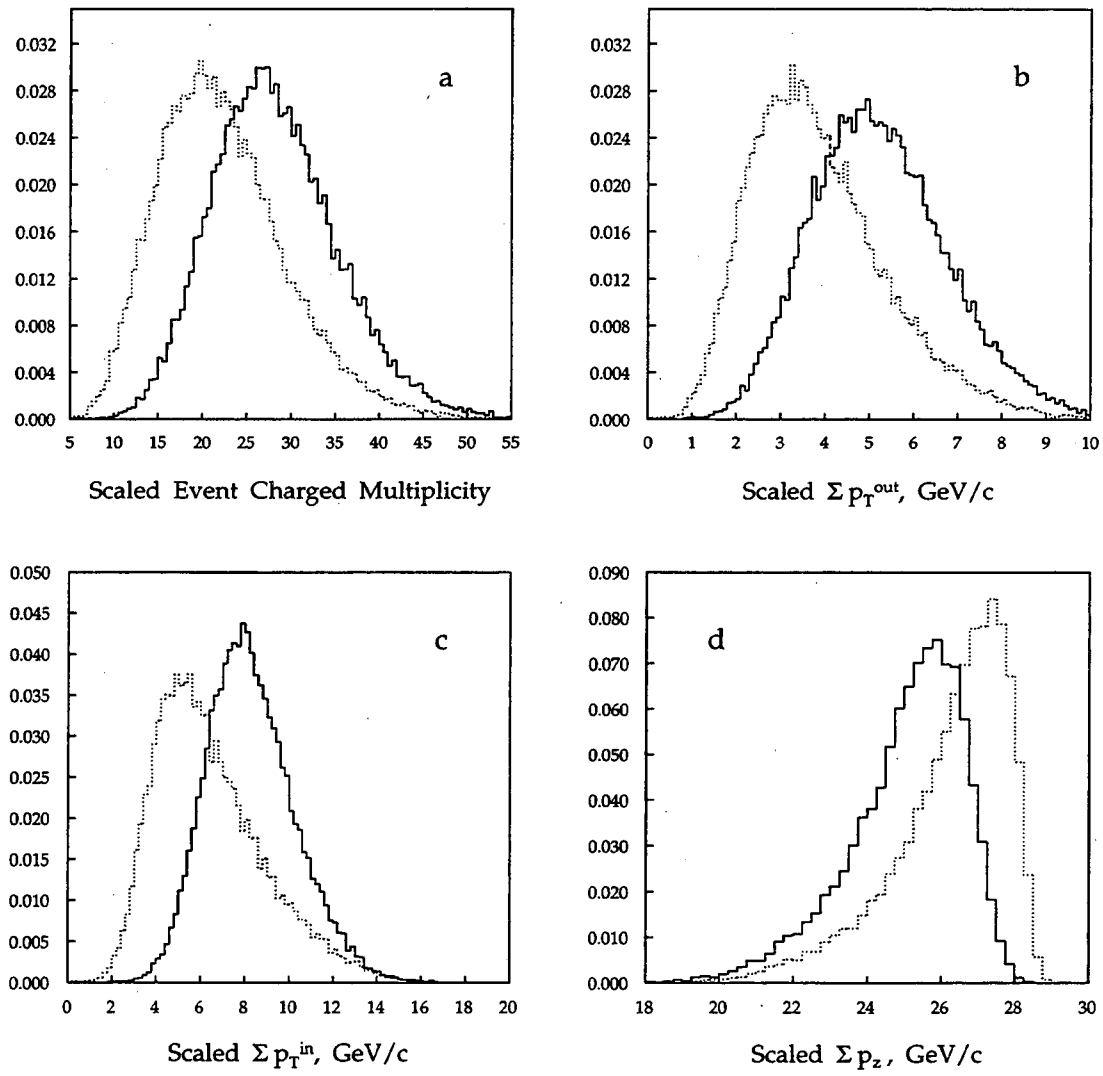


Figure 7.1: The distributions of b events (solid) and non-b events (dotted) for the inputs to the event-tagging neural network: scaled N (a), scaled $\sum p_{\perp}^{out}$ (b), scaled $\sum p_{\perp}^{in}$ (c), scaled $\sum p_{\parallel}$ (d). The area under each curve is 1.

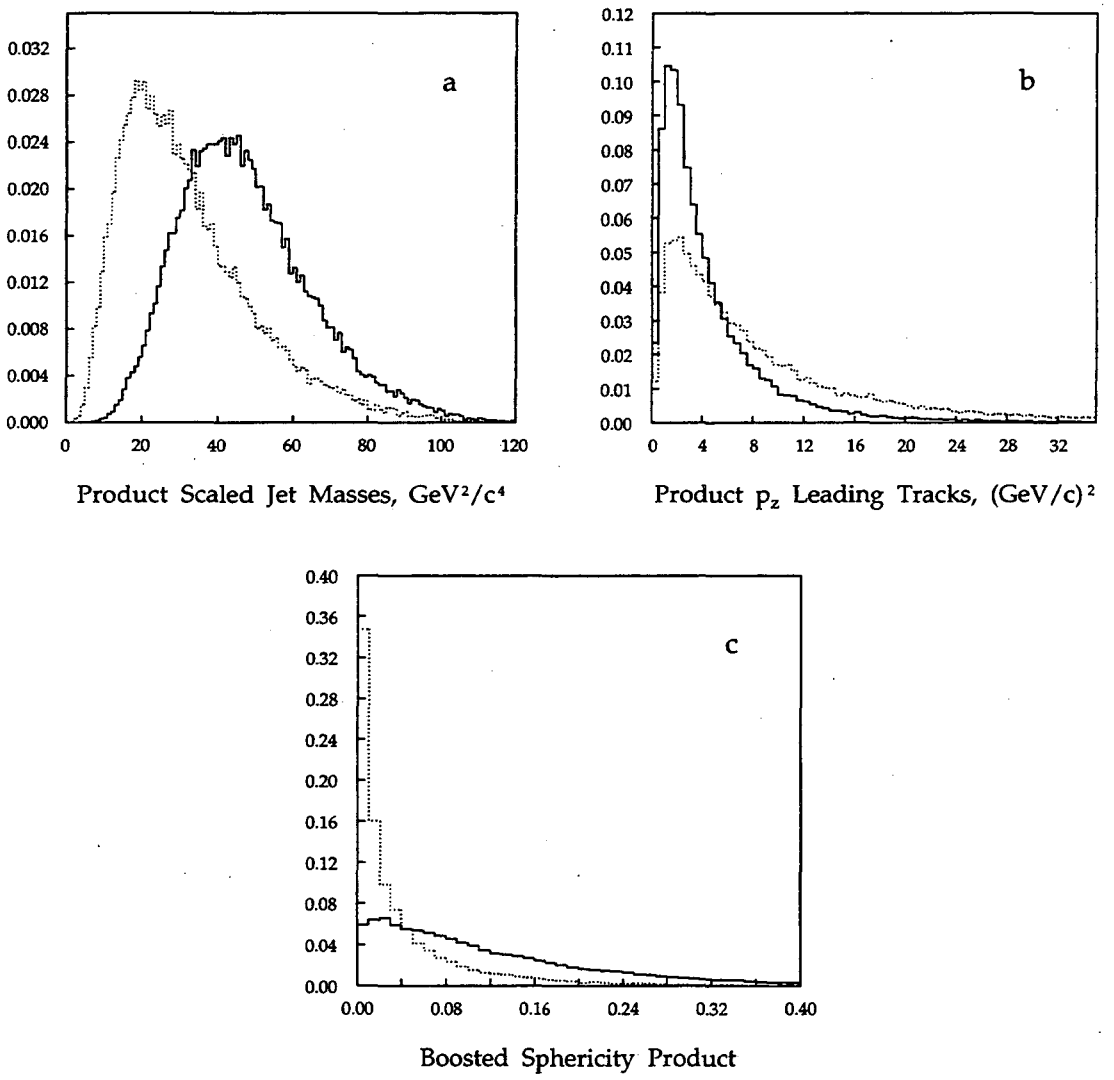


Figure 7.2: The distributions of b events (solid) and non-b events (dotted) for the inputs to the event-tagging neural network: the product of the scaled invariant masses of both hemispheres (a), $p_{||1}^{(1)} \times p_{||1}^{(2)}$ (b), and the boosted sphericity product (c). The area under each curve is 1.

The network was trained on patterns (sets of inputs) calculated from Monte Carlo events. Training was done using 7400 training patterns, half bottom and half non-bottom, following the Jetnet 2.0 default: the network was presented with a randomly selected set of 10 patterns, 5 from b events and 5 from non-b events, and the network parameters were updated after each of these sets of 10 patterns were presented. An epoch was defined to be the presentation of a set of training patterns equal in size to the entire set of training patterns. After each epoch, an independent set of 20000 test patterns made up of 50% b patterns and 50% non-b patterns was presented to the network to evaluate F. We show in Figure 7.5 how F for the training and test sets varied as a function of the epoch number during the training of the network used for the analysis presented in this chapter. During training, a running average of F for the current epoch and the previous nine epochs was calculated. Training was terminated at epoch number 4737, when this running average reached a maximum.

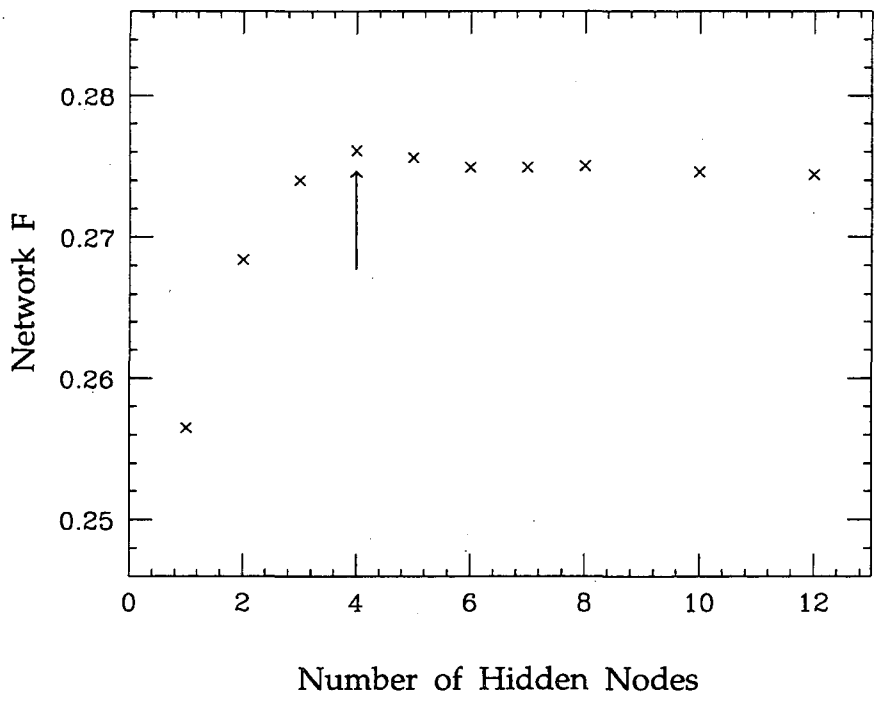


Figure 7.3: F as a function of the number of hidden nodes for the Event-Tagging Neural Network. The arrow marks the chosen number of hidden nodes (4).

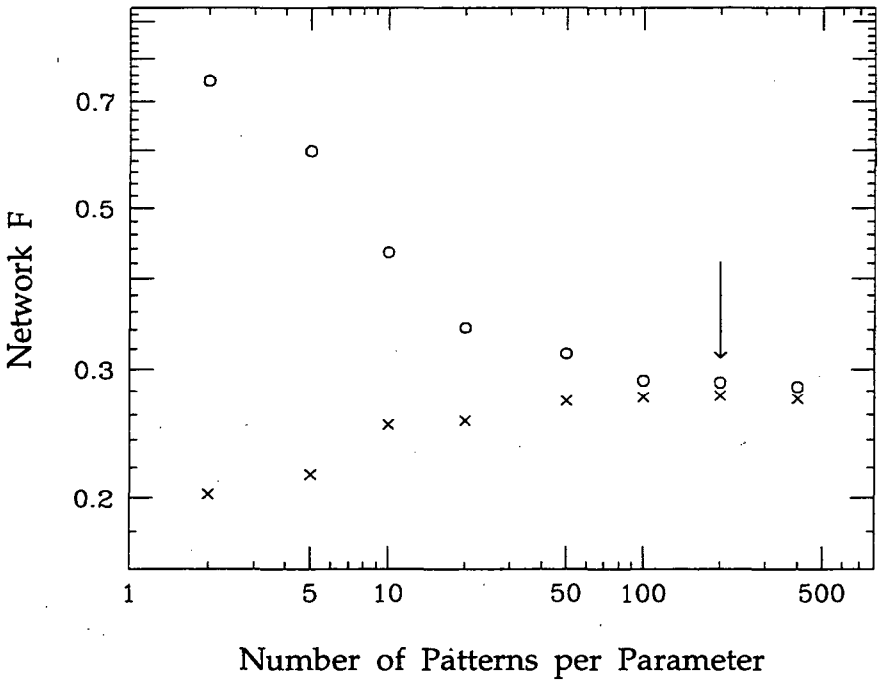


Figure 7.4: Training set F (circles) and test set F (crosses) as a function of the number of patterns per Event-Tagging Network parameter. The arrow marks the chosen number of patterns per parameter (200).

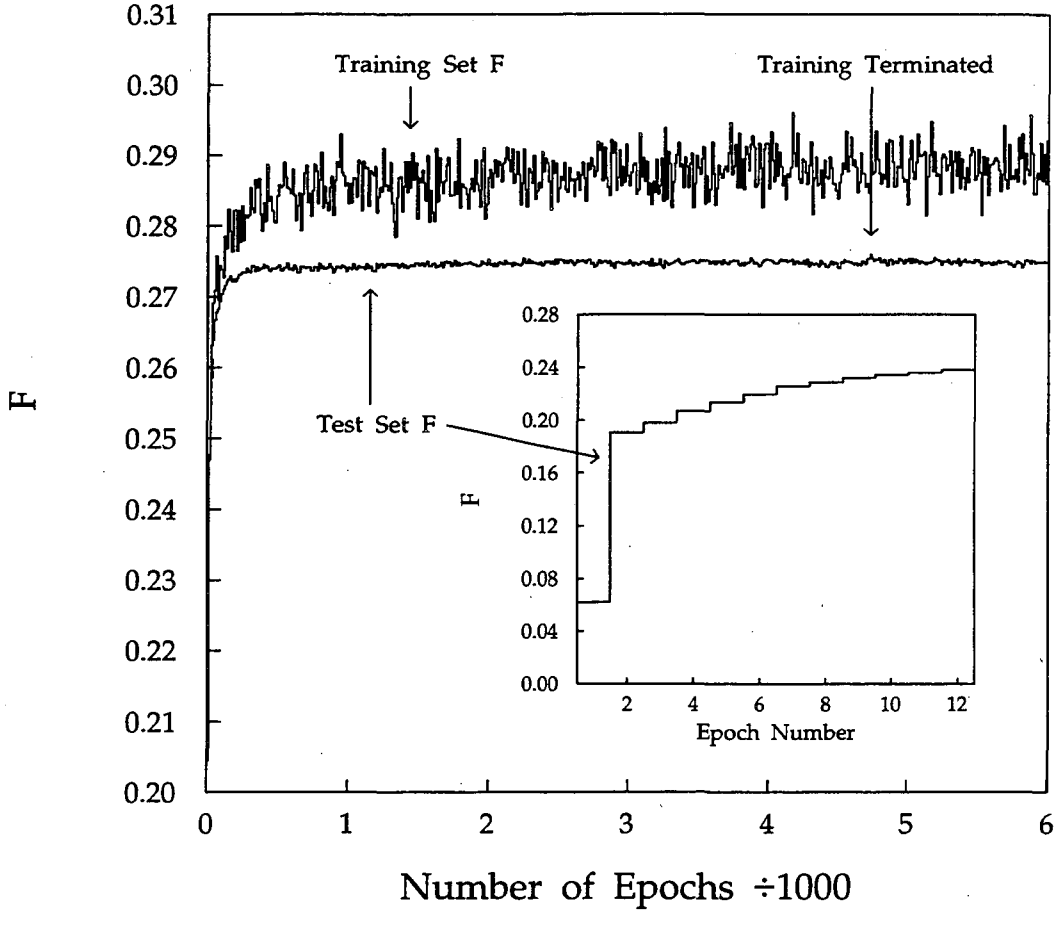


Figure 7.5: F as a function of epoch number for the Event-Tagging Neural Network. Inset: F for the test set for the first 20 epochs.

7.3 The Method for Fitting the Bottom Event Fraction

We use the trained network to create a histogram of the number of events as a function of the network output for Experiment 14-18 events that pass the analysis selections (Figure 7.6). We also use the trained network to create two histograms of the number of events as a function of the network output for an independent set of Monte Carlo events that pass the analysis selections, one histogram for bottom events and one for non-bottom events (Figure 7.7).

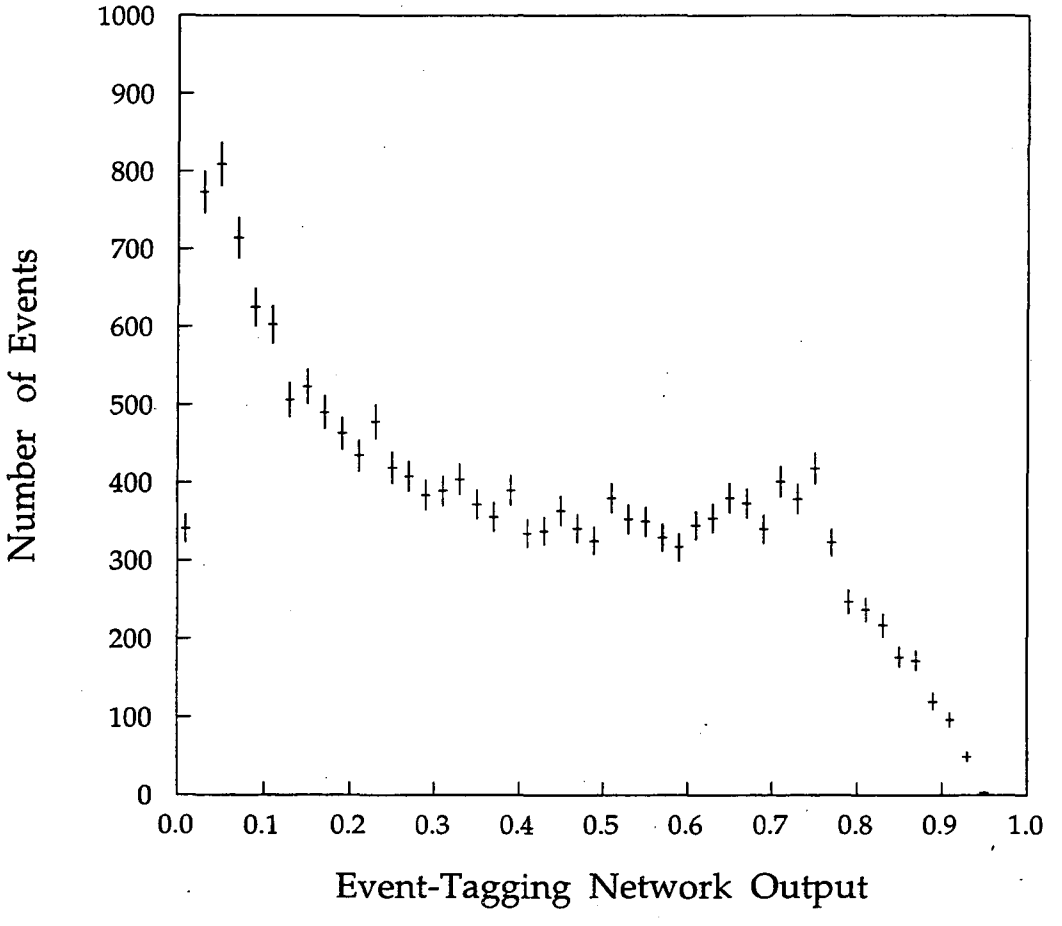


Figure 7.6: The event-tagging neural network output distributions for Experiment 14-18 data. The errors are statistical only.

We then fit a linear combination of the two Monte Carlo histograms to the histogram of experimental data. The fact that the contents of all the bins in all three histograms are subject to Poisson statistical fluctuations complicates the fit. We use a fit method, described below, that takes into account the Poisson

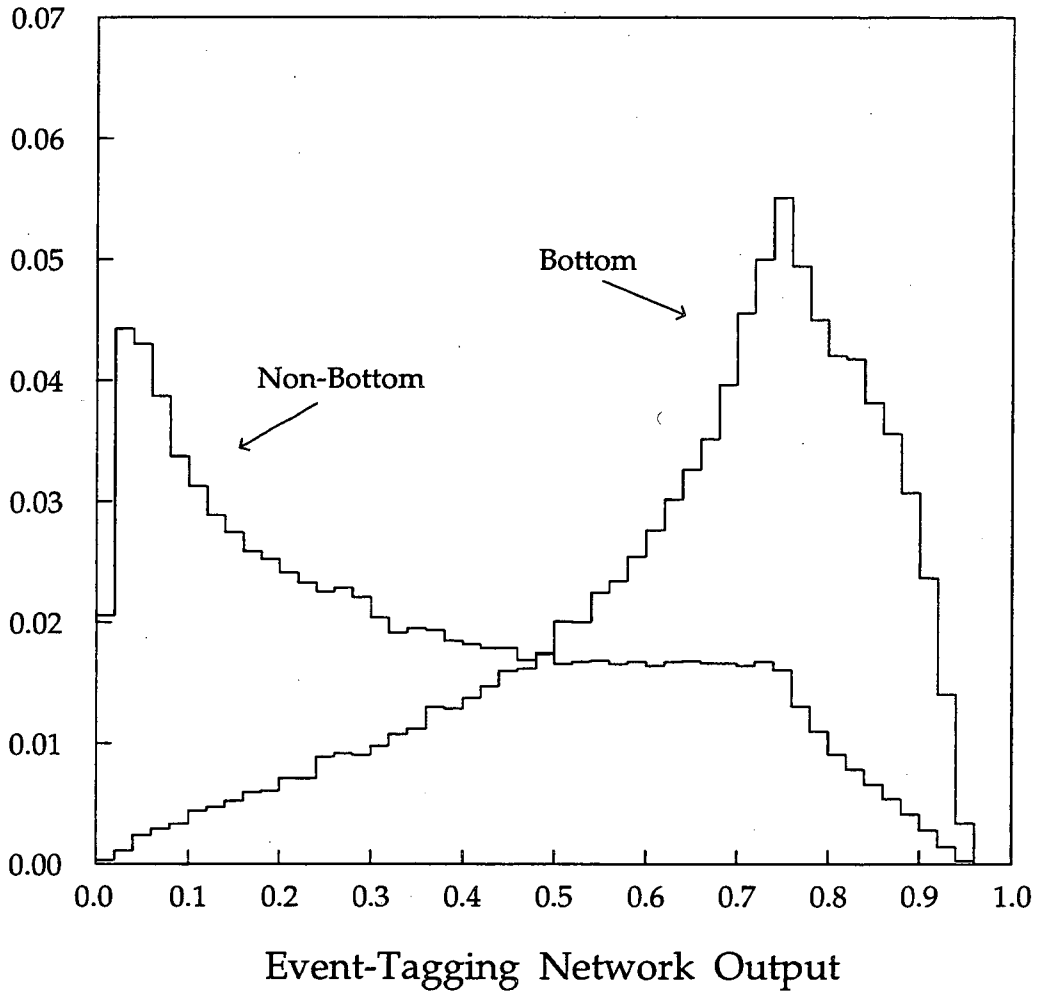


Figure 7.7: The event-tagging neural network output distributions for Monte Carlo b and non-b events. The area under each curve is 1.

statistical fluctuations in the two Monte Carlo histograms and in the experimental data histogram.

7.3.1 The Extended Maximum Likelihood Method

The fit method we use is an application of the Extended Maximum Likelihood Method, which we now review. Suppose we measure some properties $\vec{x}$ of N independent events to be $\vec{x}_k$, labeling the events by the index $k = 1, 2, ...N$. If the normalized probability density distribution of an event having a result $\vec{x}$ is $P(\vec{x}, \vec{a})$, where $\vec{a}$ are a set of unknown parameters to be determined, then the *likelihood* of measuring $\vec{x}_k$ is

$$\mathcal{L} = \prod_{k=1}^{N} P(\vec{x}_k, \vec{a}) \ . \tag{7.1}$$

The most likely values of the unknown parameters $\vec{a}$ are those that maximize $\mathcal{L}$. It is equivalent to maximize the log likelihood

$$\ell = \ln(\mathcal{L}) = \sum_{k=1}^{N} \ln\left[P(\vec{x}_k, \vec{a})\right] \ , \tag{7.2}$$

since the logarithm is a monotonic function.

Now suppose we let the observed number of events N itself be a random variable. We have assumed that the events are independent of each other, so N is Poisson distributed, with expectation value ϕ. Then the *extended likelihood* for measuring N events with their measured properties is

$$\mathcal{L}_E = \frac{\exp(-\phi)\phi^N}{N!} \prod_{k=1}^{N} P(\vec{x}_k, \vec{a}) \ . \tag{7.3}$$

The $N!$ is commonly dropped, since it is fixed and can not influence the maximum of $\mathcal{L}_E$.

7.3.2 The Event Fraction Likelihood Function

In our application of the Extended Maximum Likelihood Method [117], we let n_j be the number of entries in bin j of the data distribution being fit to, and let the expectation value of n_j be λ_j. Also, let the number of entries in bin j of Monte Carlo distribution i be m_{ij}, with expectation value f_{ij}. In fitting a linear combination of the Monte Carlo distributions to a data distribution, we make the identification

$$\lambda_j = \sum_i a_i f_{ij} \tag{7.4}$$

for all bins j, where the non-negative numbers a_i quantify how much process i contributes to the data.

We can find the precise meaning of a_i. Letting $\mathcal{F}_{ij}$ be the fraction of entries in bin j that were produced by process i, and dividing Equation 7.4 through by λ_j, we get

$$1 = \frac{1}{\lambda_j} \sum_i a_i f_{ij} = \sum_i \mathcal{F}_{ij} \ , \tag{7.5}$$

implying that

$$\mathcal{F}_{ij} = \frac{1}{\lambda_j} \ a_i f_{ij} \ . \tag{7.6}$$

The extended likelihood for the fit of a linear combination of the Monte Carlo distributions to a data distribution is the product of the Poisson probabilities to find the observed number of entries in each bin in all the histograms:

$$\begin{aligned} \mathcal{L}_E &= \left[\prod_j \frac{\exp(-\lambda_j)\lambda_j^{n_j}}{n_j!}\right] \left[\prod_{ij} \frac{\exp(-f_{ij}) f_{ij}^{m_{ij}}}{m_{ij}!}\right] \\ &= \prod_j \left[\frac{\exp(-\lambda_j)\lambda_j^{n_j}}{n_j!} \prod_i \frac{\exp(-f_{ij}) f_{ij}^{m_{ij}}}{m_{ij}!}\right] . \end{aligned} \tag{7.7}$$

Dropping the constants $n_j!$ and $m_{ij}!$, the log likelihood is

$$\ell_E = \sum_j \left[(n_j \ln \lambda_j - \lambda_j) + \sum_i (m_{ij} \ln f_{ij} - f_{ij}) \right] \ . \tag{7.8}$$

This can be rewritten as

$$\ell_E = \sum_j (\ell_E)_j \ , \tag{7.9}$$

with

$$(\ell_E)_j = (n_j \ln \lambda_j - \lambda_j) + \sum_i (m_{ij} \ln f_{ij} - f_{ij}) \ . \tag{7.10}$$

We call this likelihood the *event fraction likelihood function.*

Notice that the log likelihood is the sum over j of a function of f_{ij} and λ_j. This means the bins are independent of each other, as long as the a_i are held fixed. Therefore, the maximization of ℓ_E can be greatly simplified. Instead of having to maximize ℓ_E with respect to all parameters a_i, f_{ij}, and λ_j all at once, we can maximize with respect to only the a_i using MINUIT [118], and at each step in the

values of the a_i we can exploit the independence of each bin from all other bins and break ℓ_E into the sum over j of $(\ell_E)_j$ and optimize each bin's $(\ell_E)_j$ with respect to the f_{ij} and λ_j in that same bin.

Define u_{ij} to be

$$u_{ij} = \frac{\partial(\ell_E)_j}{\partial f_{ij}} = \left(\frac{n_j}{\lambda_j} - 1\right) a_i + \frac{m_{ij}}{f_{ij}} - 1 \,. \tag{7.11}$$

At the maximum of $(\ell_E)_j$, for a single value of j, either $u_{ij} = 0$, or f_{ij} must be at the edge of the physical region defined by $f_{ij} \geq 0$. If $f_{ij} = 0$, then $u_{ij} < 0$ is required. The optimization of $(\ell_E)_j$ must be broken into a number of cases. We derive the optimization of $(\ell_E)_j$ here for only the most common case. For the other cases, we only show the values of the f_{ij} and λ_j that optimize $(\ell_E)_j$, and give the proof that these values optimize $(\ell_E)_j$ in Appendix A.

$m_{ij} \neq 0$ for all i: In this case, if $f_{ij} = 0$ for any i, then Equation 7.10 implies $(\ell_E)_j = -\infty$. This is clearly not a maximum, so $f_{ij} \neq 0$ and $u_{ij} = 0$ for all i. Setting $u_{ij} = 0$ in Equation 7.11 and solving for f_{ij},

$$f_{ij} = \frac{m_{ij}}{1 - a_i\left(\frac{n_j}{\lambda_j} - 1\right)} \,. \tag{7.12}$$

Note that in order for m_{ij} and f_{ij} to be both positive, Equation 7.12 requires that

$$\frac{n_j}{\lambda_j} - 1 < \frac{1}{a_i} \,. \tag{7.13}$$

Multiplying Equation 7.12 through by a_i and summing over i, we get

$$\sum_i a_i f_{ij} = \lambda_j = \sum_i \frac{a_i m_{ij}}{1 - a_i\left(\frac{n_j}{\lambda_j} - 1\right)} \,, \tag{7.14}$$

which is one equation with the one unknown λ_j. Once λ_j is solved for, we can use it and Equation 7.12 to solve for f_{ij}. An efficient way of solving for λ_j is described elsewhere [117].

Some but not all $m_{ij} = 0$: In this case, a trial solution is formed. Assume that $f_{ij} = 0$ for those values of i for which $m_{ij} = 0$, use Equation 7.14 to produce an equation for λ_j, solve for λ_j using the method described in Reference [117], and use Equation 7.12 to solve for the other f_{ij}. Let $a_{M'}$ be the largest of the a_i's. If

$$\frac{n_j}{\lambda_j} - 1 < \frac{1}{a_{M'}} \tag{7.15}$$

is true, then the trial solution is the actual solution. Otherwise, the solution is obtained by letting

$$\lambda_j = n_j \frac{a_{M'}}{1 + a_{M'}} \tag{7.16}$$

using Equation 7.12 to solve for f_{ij} for $i \neq M'$, and setting

$$f_{M'j} = \frac{\lambda_j - \sum_{i \neq M'} a_i f_{ij}}{a_{M'}} \ . \tag{7.17}$$

$\boldsymbol{m_{ij} = 0}$ **for all** $\boldsymbol{i}$**:** The solution is all f_{ij} are zero, except for $i = M'$ where

$$f_{M'j} = \frac{n_j}{1 + a_{M'}} \ , \tag{7.18}$$

and

$$\lambda_j = a_{M'} f_{M'j} = n_j \frac{a_{M'}}{1 + a_{M'}} \ . \tag{7.19}$$

7.4 The Bottom Event Fraction

7.4.1 The Fit of the Bottom Event Fraction

We apply this method to the histograms in Figures 7.7 and 7.6. The fit shows that the fraction of events that pass the event selections that are b events is $10.99 \pm 0.70\%$. The error is statistical only. The fit is excellent, the χ^2 of the fit is 37.0 with 47 degrees of freedom. Figure 7.8 shows the fitted linear combination of the two Monte Carlo histograms superimposed on the experimental data histogram, along with the fitted b and non-b components. Figure 7.9 shows the bottom event efficiency and purity for the sample of events above a cut on the neural network output. The efficiencies and purities are calculated using the fitted Monte Carlo b and non-b components of the experimental data shown in Figure 7.8.

7.4.2 Correcting for Backgrounds

The fitted b event fraction must be corrected for the presence of background events in the experimental data set being fitted to.

Our Monte Carlo estimates that 0.021%, or 3.8, of the events that pass the event selections are τ events. We assign a conservative systematic error of 100%.

We do not simulate 2γ events, but we can estimate what fraction of the events that pass our event selections are 2γ events from the information used in Section 5.3.3.4. Before the τ-2γ selection was used, the excess of events in data over

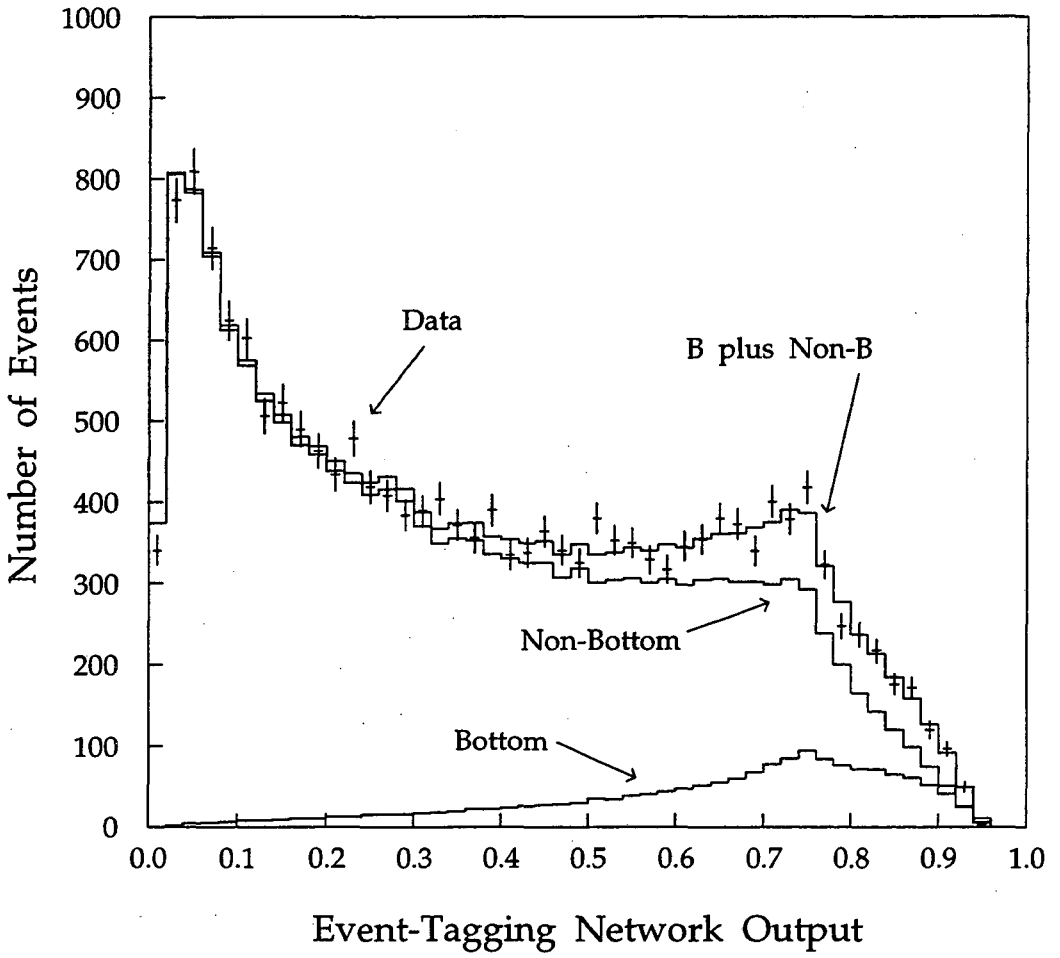


Figure 7.8: The fitted event-tagging neural network output distributions for Monte Carlo b and non-b events, their sum, and the output distribution for Experiment 14-18 events.

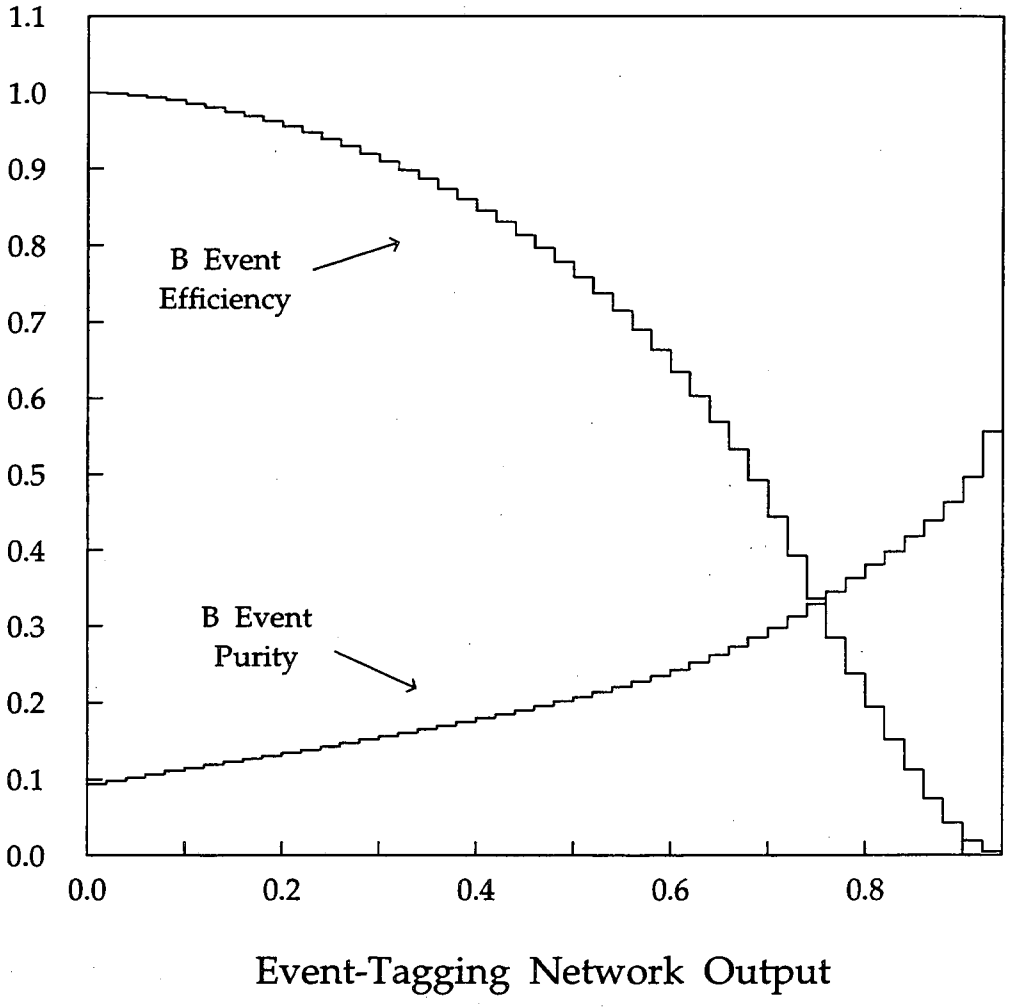


Figure 7.9: Bottom event efficiency and purity for the sample of events with network output above the cut on the abscissa.

Monte Carlo for charged multiplicities 5 and 6 was 142 ± 30 of the 25191 events in Figure 5.1a; after the τ-2γ selection was introduced this excess of τ and 2γ events becomes -1 ± 15 out of the 17943 events used in the analysis. Of course, we can not have a negative number of events, so we say that this excess contains 0 ± 15 events.

According to Monte Carlo, this excess for charged multiplicities 5 and 6 contains 81% of the τ events that pass the τ-2γ event selection, the remainder having larger multiplicity. 2γ events also have small charged multiplicities, but we have not simulated 2γ events, so we can not be more specific. In light of our knowledge and our ignorance, we assume that four τ and 2γ events pass the τ-2γ selection, and we assume that the uncertainty on this number is 30, double the size of the error of the excess in Figure 5.1a. 4 ± 30 events is $0.02 \pm 0.17\%$ of the events used in the fit of the bottom event fraction.

A previous study [25] gave a 68% confidence level upper limit of 0.01% on the fraction of Bhabha events in the events that pass the Good Hadronic Event selection. Another previous study [89] set 90% confidence level upper limits on the contribution, to the sample of events that pass the Good Hadronic Event selection, of 0.02% on $e^+e^- \rightarrow \gamma\gamma(\gamma)$ events, 0.05% on $e^+e^- \rightarrow e^+e^-\ e^+e^-$events, and 0.1% on beam gas events. The contribution of $e^+e^- \rightarrow e^+e^-\ \tau^+\tau^-$ events was estimated to be $0.02 \pm 0.01\%$. Considering the efficiency with which our additional τ-2γ selection remove τ and 2γ events that pass the Good Hadronic Event selection, it is safe to assume that none of these backgrounds are present in the events used in this analysis.

The τ and 2γ events that pass the event selections are all at small values of the neural network output, so the fit will lump them with non-bottom events. Therefore, of all the events that we fit, $10.99 \pm 0.70\%$ are bottom, $0.02 \pm 0.17\%$ are background, and $100\% - (10.99 \pm 0.70\%) - (0.02 \pm 0.17\%) = 88.99 \pm 0.72\%$ are non-bottom $q\bar{q}$ events. We conclude that $(10.99 \pm 0.70\%)/(100 - 0.02 \pm 0.17) = 11.01 \pm 0.72\%$ of the $q\bar{q}$ events are bottom events.

7.4.3 Acceptance and Physics Corrections

We now correct the fitted b event fraction for the acceptance of the TPC and the event selections. We also correct for all of the physics effects discussed in Section 2.7, and we report the ratio $\sigma(b\bar{b})/\sigma(q\bar{q})$ in QED with zero-mass quarks (which we call *massless QED*).

Let N_B^{fit} and N_L^{fit} be the number of bottom and non-bottom events that we find in the fit. Also, let N_B^{QED} and N_L^{QED} be the corresponding number of bottom and non-bottom events in massless QED. Define $C_B = N_B^{QED}/N_B^{fit}$ and $C_L = N_L^{QED}/N_L^{fit}$. We fit the bottom event fraction $f^{fit} = N_B^{fit}/(N_B^{fit} + N_L^{fit})$, then the

corresponding fraction in massless QED that we measure is

$$\begin{aligned} f_{meas}^{QED} &= \frac{N_B^{QED}}{N_B^{QED} + N_L^{QED}} \\ &= \frac{C_B N_B^{fit}}{C_B N_B^{fit} + C_L N_L^{fit}} \\ &= \left[1 + \frac{C_L N_L^{fit}}{C_B N_B^{fit}}\right]^{-1} \\ &= \left[1 + \frac{C_L}{C_B} \frac{N_L^{fit} + N_B^{fit} - N_B^{fit}}{N_B^{fit}}\right]^{-1} \\ &= \left[1 + \frac{C_L}{C_B}\left(\frac{1}{f^{fit}} - 1\right)\right]^{-1} . \end{aligned} \tag{7.20}$$

We estimate, from Monte Carlo, that $C_L/C_B = 1.166 \pm 0.008$. Equation 7.20 then implies that $f_{meas}^{QED} = 9.58 \pm 0.62\%$.

It is important to note that this value of f_{meas}^{QED} is obtained using a Monte Carlo that assumes that the bottom quark event production fraction, before physics corrections, is $1/11 \approx 9.09\%$. Therefore, we shall use the notation $f_{meas}^{QED}(9.09) = 9.58 \pm 0.62\%$. We shall deal with this dependence of f_{meas}^{QED} upon the Monte Carlo bottom quark event production fraction in Section 7.6.

7.5 The Evaluation of Systematic Errors

In our measurement, we have relied upon a Monte Carlo simulation of the TPC detector and of the process $e^+e^- \rightarrow q\bar{q} \rightarrow$ *hadrons*. To take into account this reliance, we assign a systematic error to the b event fraction that we measure. This error, which we evaluate in this section, quantifies the possible variation in the measured b event fraction due to reasonable modifications of the Monte Carlo.

There are two types of sources of systematic error. Uncertainties in the simulation of $e^+e^- \rightarrow q\bar{q} \rightarrow$ *hadrons* that produce systematic errors are:

1. The uncertainties in the values of ϵ_b and ϵ_c, as listed in Table 5.3.

2. The simulation of the decays of charm and bottom hadrons, as described in Section 5.3.1.

3. The tuning of the Lund event shape parameters a, b, Λ_{QCD}, σ_q, and Q_0, described in Section 5.3.3.

4. The uncertainty in the mass of the bottom quark.

5. The tuning of the Lund flavor parameters listed in Table 5.1.

Uncertainties in the detector simulation that produce systematic errors are:

1. The uncertainty in the number of radiation lengths in front of the TPC ($\pm 15\%$).
2. The uncertainty in the simulation of nuclear interactions of particles in the detector material in front of the TPC.
3. The uncertainty in the tracking pattern recognition efficiency (The efficiency is $97\% \pm 2\%$).

In order to find a particular systematic error, we make a variation in the Monte Carlo that reflects the estimated uncertainty in the Monte Carlo. Using this new Monte Carlo set-up, we then repeat the entire process of extracting the bottom event fraction from the experimental data, as discussed prior to this section:

1. When necessary, we retune the Jetset event shape parameters as described in Section 5.3.3. We dicuss below how we decide that retuning is necessary.
2. We train a neural network, as described in Section 7.2, on patterns calculated from Monte Carlo events that were generated using the new Monte Carlo set-up.
3. We use this new neural network to produce three new histograms of the number of events as a function of the neural network output, one for Experiment 14-18 events, one for an independent set of Monte Carlo bottom events, and one for an independent set of non-bottom events.
4. We fit a linear combination of the two new Monte Carlo histograms to the new Experiment 14-18 histogram, using the method of Section 7.3.1.
5. We correct the new fitted fraction of bottom events for backgrounds (Section 7.4.2), and for detector acceptance and physics (Section 7.4.3), obtaining the bottom fraction in QED with massless fermions (f_{meas}^{QED}).

The systematic error is the absolute value of the difference between this new value of f_{meas}^{QED} and 9.58%, the value of $f_{meas}^{QED}(9.09)$ mentioned at the end of the last section. Note that since we repeat the entire process of measuring the bottom event fraction, finding a new acceptance in the process, we do not estimate a separate systematic error on the acceptance of the event selections.

In evaluating systematic errors, there are one or two steps omitted in extracting the bottom event fraction from the experimental data. One step we always omit

is the optimization of the number of hidden nodes and the number of training patterns per network parameter, since the change in network performance is small. We also omit the tuning of the Lund event shape parameters, but if the resulting systematic error is large, this means that the variation made in the Monte Carlo has significantly spoiled the original tuning of the shape parameters, so we redo the tuning and re-evaluate the systematic error. A retune of the Monte Carlo was done to obtain the systematic errors on ϵ_b, ϵ_c, and the flavor parameter tune. The estimate of a systematic error is larger with a retune of the event shape parameters than the estimate without a retune. Therefore skipping the retune leads to a overestimate of the systematic error. The overestimate of the total systematic error is at most 35%. All of the tunes of the shape parameters we use in this analysis are listed in Table 7.2.

When the variation in the Monte Carlo is in one parameter, that parameter is varied by 2 standard deviations, in order to be conservative. Each variation in the Monte Carlo is in one direction, not both, and we assume that the systematic errors are symmetric. The systematic errors we discuss in this section are listed in Table 7.3. We now go on to discuss the individual systematic errors.

7.5.1 Systematic Errors in the Simulation of $e^+e^- \rightarrow q\bar{q}$

The b and c hadron decay simulation was discussed in Section 5.3.1. We estimate the systematic error due to the dependence on this simulation by changing the bottom hadron semileptonic decay branching ratio by 2σ, from 20.2% to 22.2%. We estimate the systematic error this way because this is the single change in the bottom decay tables that has the largest effect on the multiplicity of bottom hadron decays. The systematic error due to the uncertainty in the b and c hadron decay simulation is 0.06%.

The tuning of the Lund flavor parameters was also discussed in Section 5.3.1. The systematic error due to this source is obtained by redoing the analysis with the default Lund flavor parameters, listed in Table 5.1, substituted for the tuned flavor parameters. The systematic error due to the uncertainty in the Lund flavor parameter tune is 0.45%.

The systematic error due to the tuning of the Lund event shape parameters is estimated by doing the analysis with 3 other tunes, and adding the changes in the measured b event fraction in quadrature. Two of these tunes are simply different tunes with a 2% error floor, the first being a tune done using the coarse tuning method of Section 5.3.3.2 with Q_0 fixed at 1.0, and the second a tune using the fine tuning method with Q_0 fixed at 1.0. The third tune is the tune in Table 5.5 with a 0% error floor. The event shape parameters for these three tunes are listed in Table 7.2. The systematic error due to the uncertainty in the event shape parameter tune is 0.30%.

The 2σ uncertainties in the values of ϵ_b and ϵ_c are listed in Table 5.3. To evaluate the systematic error due to ϵ_b, we change ϵ_b by 2σ, from 0.039 to 0.15, giving a systematic error of 0.46%, the largest systematic error in this analysis. To evaluate the systematic error due to ϵ_c, we change ϵ_c by 2σ, from 0.072 to 0.26, giving a systematic error of 0.09%.

To evaluate the systematic error due to the uncertainty in the b quark mass, we change this mass in the Monte Carlo from 5.0 GeV/c^2 to 5.5 GeV/c^2. This change does not effect the masses of the bottom hadrons in the Monte Carlo, which we hold constant since they are fixed by experiment, but this change of m_b does effect the hadronization of bottom events. The systematic error due to the uncertainty in the bottom quark mass is 0.43%.

Monte Carlo Change	Parameter					χ^2/bin
	Λ_{QCD}	Q_0	a	b	σ_q	
No Change	0.364	1.360	0.053	0.497	0.320	0.94
Coarse tune	0.400	1.000	0.400	1.100	0.330	1.40
Fine tune, $Q_0 = 1$	0.382	1.000	0.071	0.642	0.320	1.16
$\epsilon_b = 0.15$	0.364	1.360	0.053	0.497	0.320	0.99
$\epsilon_c = 0.26$	0.354	1.116	0.073	0.698	0.318	1.04
Lund flavor parameters	0.339	1.257	0.052	0.496	0.330	1.04
$f_{MC}^{QED} = 0.0330$	0.411	1.466	0.100	0.526	0.322	1.02
$f_{MC}^{QED} = 0.0724$	0.377	1.355	0.031	0.450	0.321	1.01
$f_{MC}^{QED} = 0.0909$, No Change	0.364	1.360	0.053	0.497	0.320	0.94
$f_{MC}^{QED} = 0.1088$	0.331	1.201	0.042	0.484	0.319	1.02
$f_{MC}^{QED} = 0.1424$	0.296	1.088	0.000	0.483	0.318	1.07

Table 7.2: Tune of the Lund event shape parameters in Jetset 7.2.

7.5.2 Systematic Errors due to the Detector Simulation

The detector systematic errors are calculated the same way the physics simulation systematics were: vary the Monte Carlo and see how much the measured bottom event fraction changes.

We estimate that the uncertainty in the nuclear interaction cross-sections is 30%. The systematic error due to this uncertainty is obtained by turning off the nuclear interaction simulation in the Monte Carlo, measuring the resulting bottom event fraction, and multiplying the change in the measured bottom event fraction by 0.3. The nuclear interaction systematic error is 0.32%.

Error source	Systematic error
ϵ_b	0.46%
Lund flavor parameter tune	0.45%
Bottom quark mass	0.43%
Lund event shape parameter tune	0.30%
ϵ_c	0.09%
b and c hadron decay simulation	0.06%
Pattern recognition efficiency	0.35%
Nuclear interactions	0.32%
Number of radiation lengths	0.15%
All systematics, added in quadrature	0.97%

Table 7.3: The systematic errors in the Monte Carlo.

The uncertainty in the number of radiation lengths in front of the TPC is taken to be 15%. The resulting systematic error is broken into two contributions: that part due to the change in the dE/dx energy loss, and the part due to photon conversions. Each contribution is determined in the same way the nuclear interaction error was determined: turn off the interaction and multiply the change in the measured bottom fraction by 15%. The 100% correlation between the two contributions is taken into account when they are combined. When photon conversions were turned off, the measured bottom event fraction grew by a factor of 1.76. The enormous size of this change is due to an oversight in the analysis: we do not cut tracks that a pairfinder identifies as coming from a conversion pair. We remedy this by redoing the analysis with this pairfinder selection included, with the photon conversions turned on and off, and multiplying the change in the measured bottom fraction by 15%. The resulting difference in the measured fraction, which we use in this analysis, is much smaller: 0.15%.

The pattern recogition efficiency is $97 \pm 2\%$. The corresponding systematic error is found by turning off the pattern recognition simulation in Monte Carlo so that all tracks are accepted and the efficiency becomes 100%, then multiplying the change in the measured bottom event fraction by $2/(100 - 97)$. This systematic error is 0.35%.

7.6 The Monte Carlo Bottom Event Fraction

This analysis is complicated by the fact that our Monte Carlo simulation of $e^+e^- \rightarrow q\bar{q} \rightarrow hadrons$ assumes a particular value of the bottom event fraction in massless QED (QED with zero-mass quarks), which we call f_{MC}^{QED}. Therefore, the measured bottom event fraction is a function of f_{MC}^{QED}, and we make this explicit by using the notation $f_{meas}^{QED}(f_{MC}^{QED})$. For the measurement presented in Section 7.4, we used the Standard Model value of $f_{MC}^{QED} = 1/11 \approx 0.0909$; this is the reason we used the notation $f_{meas}^{QED}(9.09)$ in that section.

Unless $f_{meas}^{QED}(f_{MC}^{QED})$ is independent of f_{MC}^{QED} (in this analysis, it is not), we need to know what value of f_{MC}^{QED} to use so we can report a definite value of f_{meas}^{QED} as the bottom event fraction. The bottom event fraction is a physical quantity and can not have one value in the experimental data and another in the Monte Carlo, leading us to conclude that the physical bottom event fraction is the point where the expectation value of f_{meas}^{QED} is equal to f_{MC}^{QED}. The expectation value of any quantity is accessible only in the limit of infinite statistics, so we use f_{meas}^{QED} as the estimator of its expectation value.

To find where f_{meas}^{QED} is equal to f_{MC}^{QED}, the measurement of f_{meas}^{QED} must be repeated a number of times, each time using a different value of f_{MC}^{QED}, as discussed in Sections 7.2 through 7.4.3: tuning the Monte Carlo, training the neural network,

producing histograms of the number of events as a function of the network output for experimental data and for b and non-b Monte Carlo, fitting the Monte Carlo histograms to the experimental data histogram, and correcting for backgrounds and acceptance. The network need not be re-optimized. A curve is then fit to the variation of f_{meas}^{QED} as a function of f_{MC}^{QED}, and the point on this curve that intersects the diagonal in the f_{meas}^{QED}-f_{MC}^{QED} plane, where $f_{meas}^{QED} = f_{MC}^{QED}$, is the value of the bottom event fraction we report in this measurement. We refer to the reported value of the bottom event fraction as f^{QED}.

We repeat the measurement of f_{meas}^{QED} four times, each time using a different value of f_{MC}^{QED}. The resulting five ordered pairs $\left(f_{MC}^{QED}, f_{meas}^{QED}(f_{MC}^{QED})\right)$ are listed in Table 7.4, and plotted in Figure 7.10.

f_{MC}^{QED}	f_{meas}^{QED}
0.0330	0.0688 ± 0.0061
0.0724	0.0842 ± 0.0063
0.0909	0.0958 ± 0.0060
0.1088	0.1040 ± 0.0062
0.1424	0.1276 ± 0.0064

Table 7.4: f_{meas}^{QED} for five different values of f_{MC}^{QED}.

In Figure 7.10, we also plot the best fit of a parabola to these five points and the 1-σ statistical error contours of this fit. The parameters of the parabola are listed in Table 7.5. The fit of the parabola takes into account the high correlation of the errors on the five points being fit to. This correlation is the result of each point being a fit to the same set of experimental data, and it is why the scatter in the points about the best-fit parabola is so small compared to size of the error bars. The correlation is estimated from Monte Carlo by generating 18 sets of events that have the same number of events as the experimental data, fitting the b-fraction in each set of events for each of the five values of f_{MC}^{QED}, and computing the correlation between each pair of values of f_{meas}^{QED}. The resulting correlation matrix is in Table 7.6.

The χ^2 of the fit is 3.6 for 2 degrees of freedom, and has a confidence level of 16%, so a parabola is a good representation of how f_{meas}^{QED} varies as a function of f_{MC}^{QED}, at least in this range of f_{MC}^{QED}. The fit of a line to the five points has a χ^2 of 63 for 3 degrees of freedom, clearly a poor fit.

The best-fit parabola intersects the diagonal twice. One solution is for $f^{QED} =$ 10.2%, in the interval explored in this measurement. The other solution is at

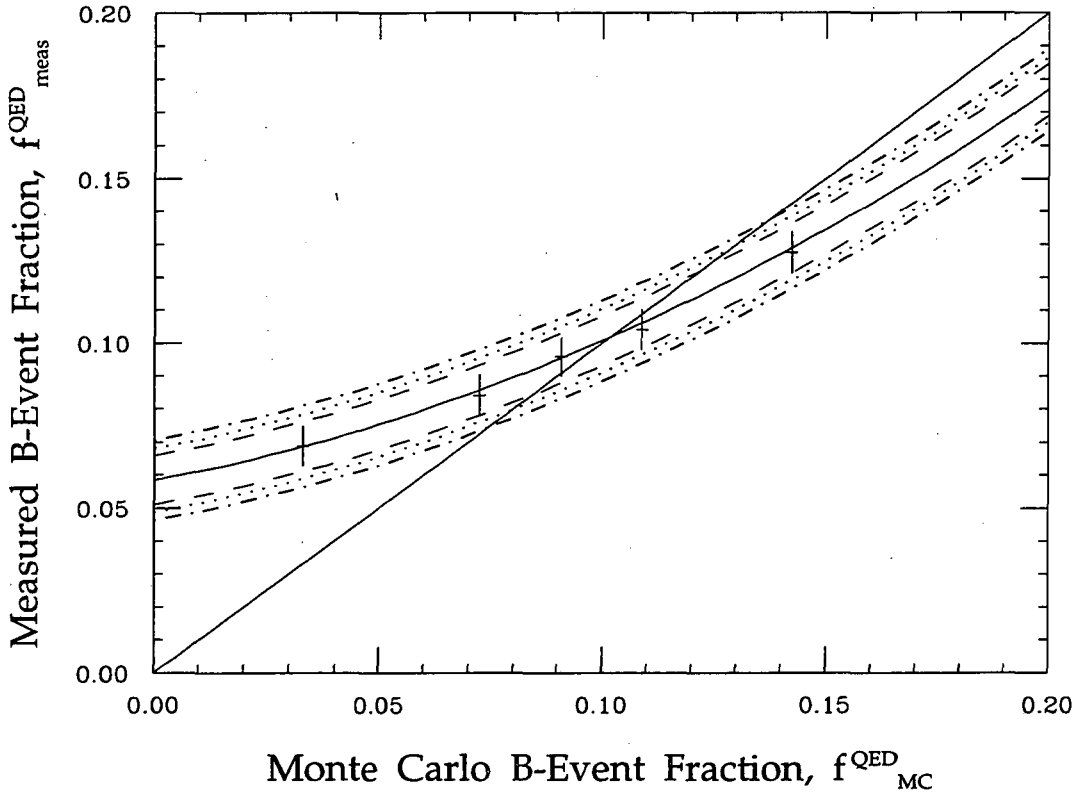


Figure 7.10: Measured b-event fraction, f_{meas}^{QED} (points with errors) as a function of the Monte Carlo b-event fraction, f_{MC}^{QED}. The best fit parabola and the diagonal ($f_{meas}^{QED} = f_{MC}^{QED}$) are solid. The statistical error 1-σ contours are the dashed curves, the systematic error 1-σ contours are the dotted curves, and the total error 1-σ contours are the dot-dash curves.

parameter	value
a_0	0.0584 ± 0.0074
a_1	0.255 ± 0.033
a_2	1.69 ± 0.22

	a_0	a_1	a_2
a_0	1	0.063	−0.11
a_1	0.063	1	−0.93
a_2	−0.11	−0.93	1

Table 7.5: The values of the parameters (left), and their correlation matrix (right), for the best-fit parabola $f_{meas}^{QED} = a_0 + a_1 f_{meas}^{QED} + a_2 (f_{meas}^{QED})^2$.

	0.0330	0.0724	0.0909	0.1088	0.1424
0.0330	1	0.987	0.994	0.976	0.971
0.0724	0.987	1	0.984	0.978	0.981
0.0909	0.994	0.984	1	0.983	0.978
0.1088	0.976	0.978	0.983	1	0.981
0.1424	0.971	0.981	0.978	0.981	1

Table 7.6: The correlation matrix for the five values of f_{meas}^{QED}.

$f^{QED} = 34.0\%$, well above the range of f_{MC}^{QED} explored in this measurement. However, Table 7.2 shows that the tuning of the shape parameters for f_{MC}^{QED} in the neighborhood of 0.1424 and above pins the Lund a parameter at $a = 0$ (we do not allow a to be negative). Therefore, we expect that how f_{meas}^{QED} varies as a function of f_{MC}^{QED} changes near $f_{MC}^{QED} = 0.1424$, and we assume that the appropriate best-fit curve for all values of f_{MC}^{QED} intersects the diagonal only once, at $f^{QED} = 10.2\%$.

The 1-σ statistical error interval on $f^{QED} = 10.2\%$ is the portion of the diagonal $f_{meas}^{QED} = f_{MC}^{QED}$ that falls between the 1-σ statistical error contours of the best-fit parabola; we obtain $f^{QED} = 10.2^{+2.0}_{-1.8}\%$. Note that this error supersedes the statistical error of $\pm 0.62\%$ on each value of f_{meas}^{QED} reported in Section 7.4.3. The absolute value of $(f_{meas}^{QED} - f_{MC}^{QED})/\sigma_{\text{stat}}(f_{meas}^{QED})$ is plotted as a function of f_{MC}^{QED} in Figure 7.11.

We need to translate the systematic and total errors on f_{meas}^{QED} into errors on f^{QED}, in the same way we have just done for the statistical error. First, we must know how the total systematic error varies as a function of f_{meas}^{QED}; the systematic error of $\pm 0.97\%$ we found in Section 7.5 was on $f_{meas}^{QED}(9.09)$. We have re-evaluated, for $f_{MC}^{QED} = 14.24\%$, the systematic errors due to the uncertainties in m_b and the pattern recognition efficiency. Both of these systematics are independent of f_{meas}^{QED}, so we can safely assume that the total systematic error is also independent of f_{meas}^{QED}. The portion of the diagonal between the 1-σ systematic error contours gives us the systematic error on f^{QED}, giving us $f^{QED} = 10.2^{+2.0+2.7}_{-1.8-2.2}\%$; the first error is statistical and the second systematic. The absolute value of $(f_{meas}^{QED} - f_{MC}^{QED})/\sigma_{\text{syst}}(f_{meas}^{QED})$ is also plotted as a function of f_{MC}^{QED} in Figure 7.11.

The portion of the diagonal between the 1-σ total error contours gives us the total error on f^{QED}: $f^{QED} = 10.2^{+3.6}_{-2.8}\%$. Note that the positive total error is *not* the statistical and systematic positive errors added in quadrature (3.4%). This is the result of the fact that f_{meas}^{QED} is not a linear function of f_{MC}^{QED}. The absolute value of $(f_{meas}^{QED} - f_{MC}^{QED})/\sigma_{\text{tot}}(f_{meas}^{QED})$ is also plotted as a function of f_{MC}^{QED} in Figure 7.11.

As a check on the method, we have generated five 160k event Monte Carlo data

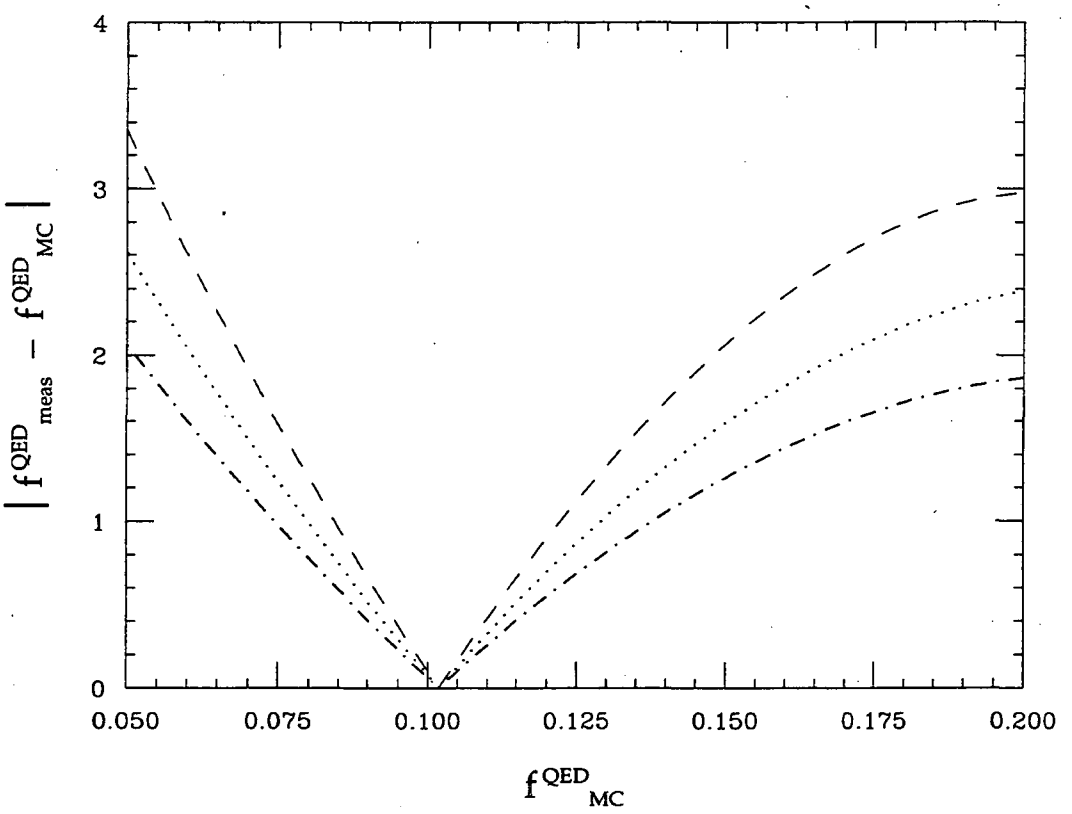


Figure 7.11: The absolute value of the difference between the measured and Monte Carlo b-event fractions, in standard deviations, as a function of the Monte Carlo b-event fraction. The dashed curve is computed using the statistical error only, the dotted curve using the systematic error only, and the dot-dash curve using the total error.

sets, each set with one of the values of f_{MC}^{QED} used in this section. We find f_{meas}^{QED} in each of these sets, for each set using the neural network and the Monte Carlo bottom and non-bottom event-tagging network output distributions created using the same value of f_{MC}^{QED}. We expect $f_{meas}^{QED} = f_{MC}^{QED}$, within statistics, for each of the five points, and this is what we find.

7.7 Discussion of the Results

This measurement of $f^{QED} = 10.2^{+2.0+2.7}_{-1.8-2.2}\% = 10.2^{+3.6}_{-2.8}\%$ is statistically compatible with the Standard Model prediction that $f^{QED} = 1/11 = 9.1\%$. The systematic error on this measurement is somewhat larger than the statistical error.

Three large systematic errors in Table 7.3 are overestimates, since the event shape parameters were not retuned. These systematic errors are due to the uncertainties in the bottom quark mass (0.43%), the pattern recognition efficiency (0.35%), and the nuclear interaction simulation (0.32%). The size of these three systematics could be significantly reduced by evaluating these systematic errors doing the retune, but even if all three errors turned out to be zero, the net systematic error would shrink only 25%, and the total error on f^{QED} would shrink only 16%.

7.7.1 How $\sigma(f^{QED})$ Depends Upon $\sigma(f_{meas}^{QED})$

Each point on the best-fit parabola has a statistical uncertainty of $\pm 0.75\%$ on f_{meas}^{QED}. Yet the average of the positive and negative statistical errors on f^{QED} we report is $\pm 1.9\%$. This magnification of the statistical error, as well as the magnification of the systematic and total errors, is produced by the non-zero slope of the parabola at the point where $f_{meas}^{QED} = f_{MC}^{QED}$.

We use Figure 7.12 for a geometrical derivation of this magnification factor for any type of error. We approximate the parabola with its tangent at $f_{meas}^{QED} = f_{MC}^{QED}$. Let the vertical distance between the 1-σ contours be $2W$, where W is the error on the ordinate of each point of the tangent. W is independent of m. The hypotenuse of the triangle in Figure 7.12 has a length equal to that portion of the diagonal in Figure 7.10 between the 1-σ contours. The projection of the hypotenuse onto the horizontal axis, $2E$, is twice the error on f^{QED}. $2E$ is also the length of each leg of the triangle in Figure 7.12. Let $2E = 2W + L$, then it is true that $L/(2E) = \tan\theta = m$, and $2W = 2E - L = 2E(1-m)$. Therefore, $E = W/(1-m)$, which is the relation that we seek.

This relation is singular for $m = 1$, but the expression is not valid for m near 1, since this situation implies that the best-fit parabola is approximately tangent to the diagonal, and that the 1-σ contour is entirely above the diagonal. Since the 1-σ

contour dips significantly below the diagonal in this analysis, we do not consider the case where m is near 1. In this analysis, $m = 0.60 \pm 0.02$. For the statistical error, we predict E to be $0.75\%/(1 - 0.60) = 1.9\%$, which is the average of the positive and negative statistical errors on this measurement. For the systematic error, we predict E to be $0.97\%/(1 - 0.60) = 2.4\%$, which is the average of the positive and negative systematic errors on this measurement. For the total error, we predict E to be $1.23\%/(1 - 0.60) = 3.1\%$, which is a little less than 3.2%, the average of the positive and negative total errors on this measurement; the difference is the result of the large size of the error coupled with the non-linearity of the dependence of f_{meas}^{QED} upon f_{MC}^{QED}.

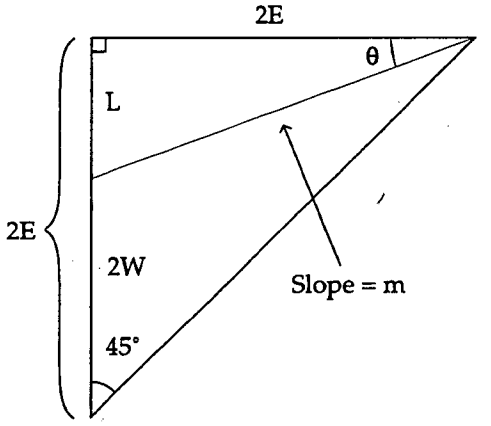


Figure 7.12: Geometrical picture of how the error is magnified.

If the variation of f_{meas}^{QED} as a function of f_{MC}^{QED} is not taken into account, as has been done here, the error on f^{QED} can be significantly underestimated. Also, the value of f_{meas}^{QED} reported can be biased if m is large enough, and if it is significantly different, statistically speaking, from the value of f_{MC}^{QED}.

7.7.2 Comparison to LEP Measurements of $\Gamma(Z^0 \to b\bar{b})$

There are three published measurements, using a neural network in a manner similar to the method used here, of the fraction of hadronic events that are bottom events. All of these are measurements of $\Gamma(Z^0 \to b\bar{b})$ at LEP [114, 115, 116]. All three fit linear combinations of histograms of the number of events as a function of the neural network output(s) for Monte Carlo events of different flavors to a histogram of the number of events as a function of the neural network output(s) for experimental data events, as described in Section 6.4.2 and References [114, 115].

In Table 7.7, we list the results of our measurement and of the three LEP

measurements. The errors for our measurement is significantly larger than that on the LEP measurements.

Each of the LEP experiments has of the order of 100 times as much data as the TPC/2γ experiment. Therefore, in the absence of other factors, we would expect the statistical error on our measurement to be 10 times the average of the statistical errors for the three LEP measurements. Table 7.7 shows that the difference is a factor of 4. The change from 10 to 4 is caused by the greater differences between the properties of bottom and non-bottom events at 29 GeV compared to 92 GeV, and by the magnification of the errors by the variation of f_{meas}^{QED} with f_{MC}^{QED} in this measurement.

Experiment	Bottom event fraction
TPC/2γ	$10.2^{+2.0+2.7}_{-1.8-2.2}\%$
L3	$22.2 \pm 0.3 \pm 0.7$ %
ALEPH	$22.8 \pm 0.5 \pm 0.5$ %
DELPHI	$23.2 \pm 0.5 \pm 1.7$ %

Table 7.7: All measurements of the bottom event fraction using a neural network. The first error is statistical, the second systematic. Our measurement of the bottom event fraction in the continuum (first row) is expected to be very different from the bottom event fraction at the Z^0 (other three rows); the purpose of this table is to compare the errors on these measured fractions, not the fractions themselves.

The central feature of this measurement is the care with which the dependence of f_{meas}^{QE} upon f_{MC}^{QED} was taken into account. There is no mention in Reference [115] (Section 6.4.2) that the L3 measurement took into account the variation of $\Gamma(Z^0 \rightarrow b\bar{b})_{meas}$ as a function of $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$. The value of $\Gamma(Z^0 \rightarrow b\bar{b})_{meas}$ found by the ALEPH measurement [114] is reported to be insensitive to large changes in $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$, which is surprising because of the sharp contrast with what we found in this measurement. It would be very interesting to know just how $\Gamma(Z^0 \rightarrow b\bar{b})_{meas}$ depends upon $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$, and to understand why the dependence of the measured bottom event fraction upon the Monte Carlo bottom event fraction is so different in the two measurements.

The DELPHI measurement [116] (Section 6.4.2) used a method of taking into account the variation of $\Gamma(Z^0 \rightarrow b\bar{b})_{meas}$ as a function of $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$ that is very different from the method used in this measurement. The DELPHI best tune was found by varying the Monte Carlo parameters to minimize the χ^2 of the difference between the distributions of data and Monte Carlo in rapidity and aplanarity. In the evaluation of systematic errors, the range of each Monte Carlo parameter was found by finding the extreme values of that parameter that make the χ^2

equal to $\chi^2_{\text{best tune}} + 16$, allowing the values of $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$ and $\Gamma(Z^0 \rightarrow c\bar{c})_{MC}$ to float to minimize the χ^2 for a fixed value of the parameter in question. This approach magnifies the systematic error, rather than the total error, in order to take into account the variation of $\Gamma(Z^0 \rightarrow b\bar{b})_{meas}$ as a function of $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$. This approach has an understandable logic to it, and perhaps it produces results similar to the approach used in this analysis. We believe that this approach has two disadvantages compared to the approach used here:

1. The DELPHI method does not show how $\Gamma(Z^0 \rightarrow b\bar{b})_{meas}$ varies as a function of $\Gamma(Z^0 \rightarrow b\bar{b})_{MC}$.

2. The correlations between Monte Carlo parameters were not taken into account in the DELPHI measurement, and this can lead to an underestimate of the range of the Monte Carlo parameters, which in turn will cause the systematic errors to be underestimated.

Chapter 8

A Study of Bias in Techniques to Measure Charged Hadron Production in Bottom Quark Jets

Up until now, neural networks have only been used in High Energy Physics, either for measurements where something is being counted, such as the number of bottom quark events in a sample, or for lower level classification problems. Both uses are described in Section 6.4.

It is highly desireable to use a neural network to do tagging measurements, especially measurements involving bottom hadrons. Measurements at LEP where identifying bottom jets with a neural network might be useful include measurements of the bottom hadron lifetime and b-mixing measurements, the latter using a network that can identify the bottom quark charge. At higher-energy colliders, tagging bottom jets is useful for identifying particles that decay predominantly into one or more bottom quarks, such as the top quark or lower-mass higgses.

In this chapter, we describe a study of bias in techniques to measure the $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$ (charged hadron) momentum spectra for bottom jets, using a neural network to identify the bottom jets. We begin this chapter with a description of our neural network for tagging bottom jets. We then show that bias is present in any measurement of charged hadron production in bottom jets, where a neural network tags the bottom quark jets. Finally, we describe an investigation of the sources of this bias.

8.1 Introduction

In any measurement technique, it is important to avoid bias as much as is practical. In some measurements, avoiding bias is impossible and a bias correction is made, but efforts are still made to design the measurement so that the bias is as small as practical. Avoiding bias is particularly important in measurements that use a neural network to tag bottom jets, since we depend heavily upon Monte Carlo to train our neural network. Avoiding bias is a factor in many of the choices made in constructing the study presented here.

One of these choices was to divide each event into two halves. The tracks in one half of the event are used for computing inputs to the neural network we use for distinguishing bottom and non-bottom events, and the tracks in the other half of the event are used to measure charged hadron production. This approach avoids using the same track to compute both the neural network inputs and the charged hadron production. To avoid bias, we also chose to use the plane perpendicular to the event thrust axis to divide the event in half, since the thrust axis tends to not have tracks close to the plane perpendicular to it [119]. We often refer to each half of an event as a jet, since we only use 2-jet events in this study (see Section 8.2), and since the tracks in each half of the event mostly come from one QCD jet.

8.2 Track and Event Selections

The track selection used in this study is the good track selection described in Section 5.2.

The event selection used are the selections described in Section 5.2, plus these additional criteria:

1. Events must be 2-jet (selected by LUCLUS with $d_{join} = 2.5$), thus eliminating obvious 3-jet events, where the gluon jet might correlate the two halves of the event and cause information to flow between the two halves of the event.

2. Events must have at least 7 charged tracks, since LUCLUS automatically classifies events with 6 or less tracks as 2-jet.

3. Events with a jet axis with dip $|\lambda| > 45°$ are cut to eliminate events with a significant number of tracks that do not enter the TPC fiducial volume. The axis of each jet is the sum of the momenta of the tracks in the jet.

4. Events where the angle between the axes of the two jets is less than 140° are cut to eliminate events with an energetic third jet that is not seen.

15091 events satisfy all of these selection criteria.

For the tracks in the tagged jet, which are used to obtain the charged hadron momentum spectra, we must apply selection criteria that are more restrictive than the good track selection described in Section 5.2:

1. Tracks must have a dip $|\lambda| < 60°$, since the track acceptance declines with increasing dip starting near 60°.

2. Tracks must have a distance of closest approach to the interaction point less than 3 cm in the x-y plane, and less than 5 cm in z, so that the track is consistent with coming from the event vertex. This criterion eliminates tracks that come from nuclear interactions or cosmic rays.

3. Either $dC < 0.15$ GeV^{-1} or $dC/C < 0.15$, where C is the curvature, and dC is the curvature error. This ensures that the track momentum is well measured.

4. There must be at least 40 wires associated with each track, since the behavior of the $\mathrm{d}E/\mathrm{d}x$ resolution is not well understood for tracks with less than 40 wires.

5. Positively charged tracks with $x_p = p/p_{\mathrm{beam}} < 0.25$ are not used, in order to exclude nuclear interaction products. The remainder of the positive tracks and all negative tracks are used in this study.

8.3 The Jet-Tagging Neural Network

Since the jet-tagging neural network's inputs contain information from only one half of each event, this network will not distinguish b from non-b as well as the event-tagging neural network described in Section 7.1. In an attempt to have good network performance, the inputs for this study's network are designed using the raw input approach described in Section 6.3.

The 15 inputs for each jet are defined as follows. A coordinate sytem is constructed as defined below, and the tracks in each jet are ordered in rapidity y ($= \frac{1}{2}\log\left[\frac{E+p_z}{E-p_z}\right]$, we assume the charged pion mass): track number 1 is the track with the largest y, track number 2 is the track with the second largest y, etc. Inputs 1-3 are p_x, p_y, and p_z (respectively) for track 1, inputs 4-6 are p_x, p_y, and p_z (respectively) for track 2, and so on for the 4 leading tracks in the jet, giving 12 inputs. Inputs 13-15 are $\sum|p_x|$, $\sum|p_y|$, and $\sum|p_z|$ (respectively), where the sum is over all but the 4 leading tracks.

We define the coordinate system two ways. The simplest way, the *uniterated* method, is illustrated in Figure 8.1. The z axis in each jet is defined to be parallel

to the thrust axis and points away from the plane that divides the event in half. The x axis is defined to be perpendicular to the event plane, which contains the thrust axis and the thrust major axis. The sign ambiguity in the choice of the perpendicular direction is handled by randomly choosing the sign. The y axis is set equal to $\hat{z} \times \hat{x}$. The x axes for the two jets are defined to point in the same direction, requiring that the y axes be in opposite directions. The coordinate systems in the two jets are fixed relative to each other, so we expect that this definition of the coordinate systems carries information between the two halves of the event. We use this definition of the coordinate system only in our study of the bias of techniques to measure charged hadron production in bottom quark jets.

The other way we define the coordinate system, the *iterated* method, is designed so that the coordinate systems for the two halves of the event are relatively independent of each other. However, any method dividing the event in half must use information from the entire event, so it is impossible for the coordinate systems in the two jets to be completely independent of each other. This method is illustrated in Figure 8.2. We use the LUCLUS cluster-finder to divide each jet into two *subjets*. $\vec{A}_1$ is equal to the sum of the momenta of the tracks in the subjet with the largest momentum, and $\vec{A}_2$ is equal to the sum of the momenta of the tracks in the other subjet. The z axis is defined to be parallel to the vector sum of the momenta of all the tracks in the jet. The x axis is $\hat{x} = (\vec{A}_1 \times \vec{A}_2)/|\vec{A}_1 \times \vec{A}_2|$. The y axis is defined to be $\hat{y} = \hat{z} \times \hat{x}$. In the iterated definition of the axes, the plane that divides the event in half generally does not contain the x and y axes, and the event plane generally does not contain $\vec{A}_1$ and $\vec{A}_2$.

We use the iterated definition in our study of the bias, and we would use it in a measurement of charged hadron production in bottom events. Figures 8.3 through 8.6 show the distributions of Monte Carlo bottom and non-bottom jets for the 15 inputs computed using the iterated definition of the axes. The distributions of the p_x and p_y inputs are asymmetric in these figures as a result of our way of defining the x and y axes. Note that many, but not all, of the inputs have different distributions for non-bottom and bottom jets. The discussion in Section 2.5 provides an explanation for the gross features of the input distributions. The p_z for the four leading tracks is smaller for bottom jets than for non-bottom jets, in agreement with the observation that hadrons in bottom jets have less momentum than hadrons in non-bottom jets. p_x and p_y for the four leading tracks are equal for bottom and non-bottom jets, in agreement with the observation that $p_\perp$ is the same, for individual tracks, for non-bottom and bottom jets. However, inputs 13 and 14 are larger for bottom jets than for non-bottom jets, since bottom jets have a larger average multiplicity, and more tracks in the sum over all but four tracks, than non-bottom jets. Input 15 is larger for bottom jets than for non-bottom jets presumably because of the larger charged multiplicity for bottom jets versus non-bottom jets.

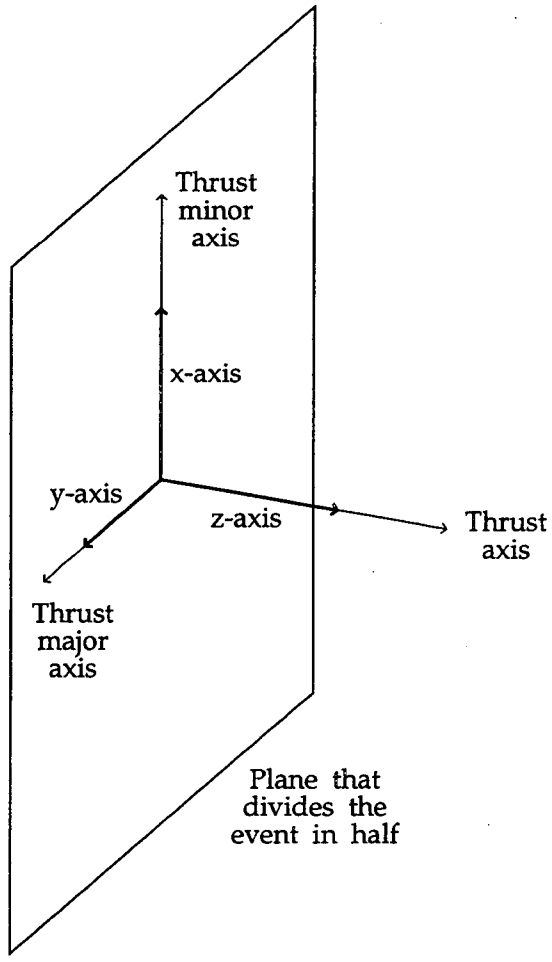


Figure 8.1: The *uniterated* coordinate system.

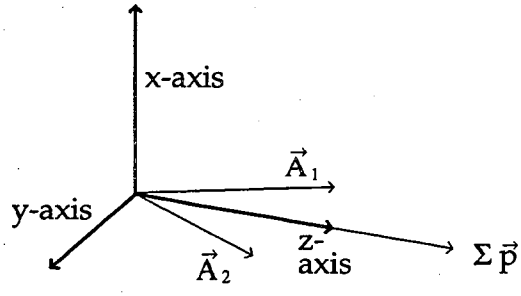


Figure 8.2: The *iterated* coordinate system. $\vec{A}_1$ is the sum of the momenta of the tracks in the subjet with the largest momentum, and $\vec{A}_2$ is the sum of the momenta of the tracks in the other subjet.

Figures 8.4 and 8.5 show that p_x and p_y for the four leading tracks have essentially identical distributions for bottom and non-bottom jets. These inputs do carry a substantial amount of information, for when these 8 inputs are eliminated and a network using the other 7 inputs is trained the same way the 15-input network is trained (described below), the network F (described in Section 6.3) declines from 0.157 to 0.109. The p_x and p_y inputs carry information that allows the reconstruction of the invariant masses of all subsets of the four leading tracks. In addition, a comparison of the relative spread of momenta in p_x versus p_y distinguishes jets in planar 3-jet events from rounder bottom events, and both from more pencil-like jets in 2-jet non-bottom events. This illustrates the ability of neural networks to distinguish the distributions of bottom and non-bottom jets from each other in the 15-dimension space of inputs, even though the projections of these distributions onto one of the inputs are sometimes identical.

A neural network with one node for each of the fifteen inputs, one hidden layer, and one output node was trained on Monte Carlo events that pass the event selections, in the same manner that the event-tagging network was trained (Section 7.2). Jets with four or fewer tracks do not have network inputs calculated, since some of the inputs would not be defined, but neither are they discarded. These low-multiplicity jets are placed in their own bin in the histograms of network outputs at $y = 1.05$ (and F is calculated taking into account this extra bin). Strictly speaking, these low-multiplicity jets do not possess a network output, and it is not possible for the network output to be 1.05, but we refer to these low-multiplicity jets as having a network output of 1.05.

The network was trained with a variety of number of hidden nodes, so that we could pick the number of hidden nodes at the lower end of the range in which the performance F levels off (Figure 8.7). The chosen number is 15, though 12 would have worked just as well. The number of patterns per network parameter was set to 200, as was done for the event-tagging network (Section 7.2). Figure 8.8 shows how F for the training and test sets varied as a function of the epoch number during the training. Figure 8.9 shows the network output distribution for Experiments 14-18, along with the extra bin at 1.05. Figure 8.10 shows the network output distributions for Monte Carlo bottom and non-bottom jets, along with the extra bin at 1.05.

We have done a fit of a linear combination of the network output distributions for Monte Carlo bottom and non-bottom jets to the output distribution for Experiment 14-18 jets, in the same way that the fit of the bottom event fraction was done in Section 7.4.1. The fitted fraction of bottom jets is 0.0747 ± 0.0079. Figure 8.11 shows the fitted linear combination of the two Monte Carlo histograms of Figure 8.10 superimposed on the experimental data histogram, along with the fitted b and non-b components. Figure 8.12 shows the bottom jet efficiency for the sample of jets above a cut on the neural network output. Also shown are the

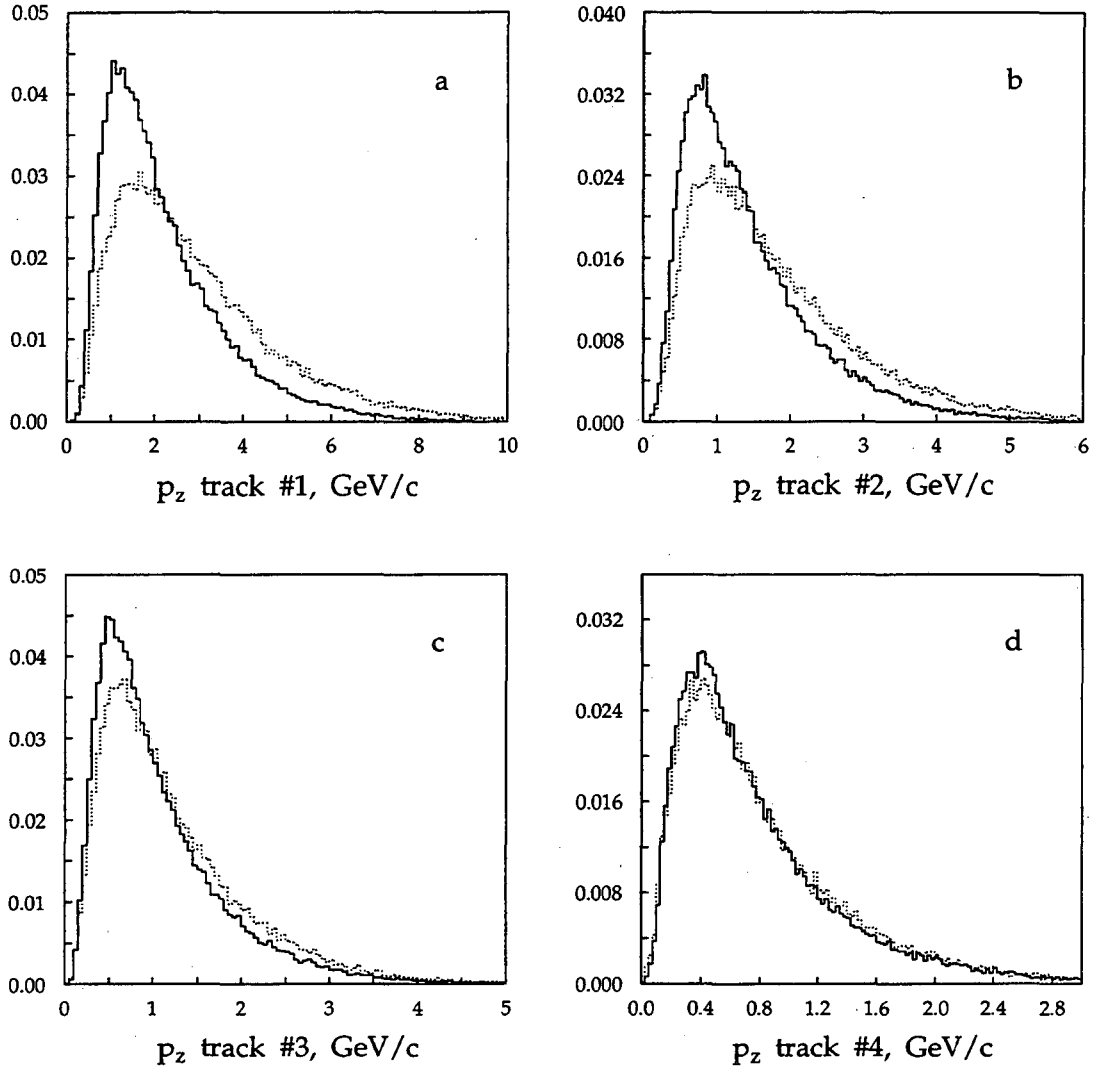


Figure 8.3: Jet-Tagging Neural Network inputs 3 (a), 6 (b), 9 (c), and 12 (d), which are the p_z for tracks 1, 2, 3, and 4, respectively. Solid is bottom, dotted is non-bottom. The area under each curve is 1.

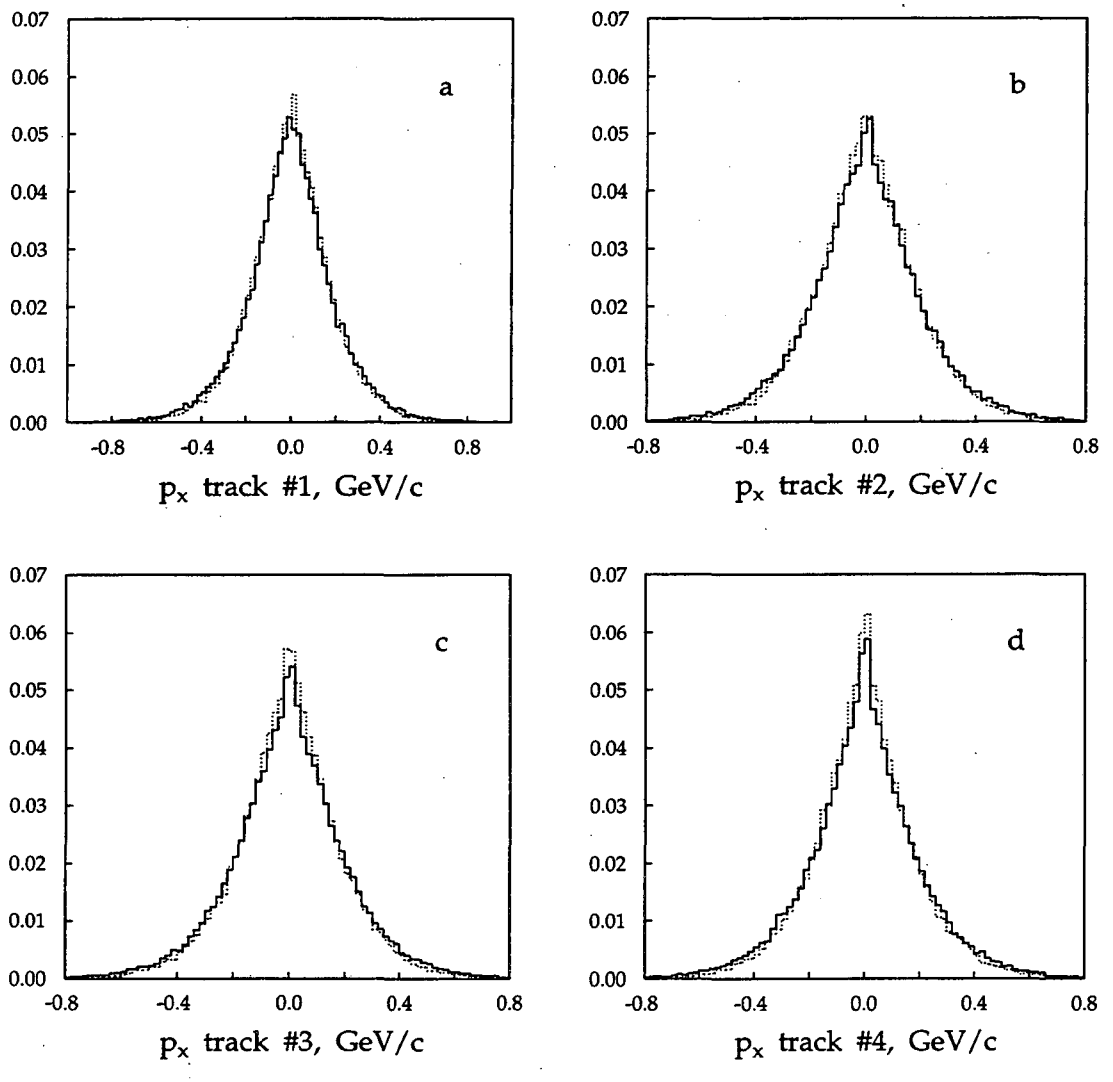


Figure 8.4: Jet-Tagging Neural Network inputs 1 (a), 4 (b), 7 (c), and 10 (d), which are the p_x for tracks 1, 2, 3, and 4, respectively. Solid is bottom, dotted is non-bottom. The area under each curve is 1.

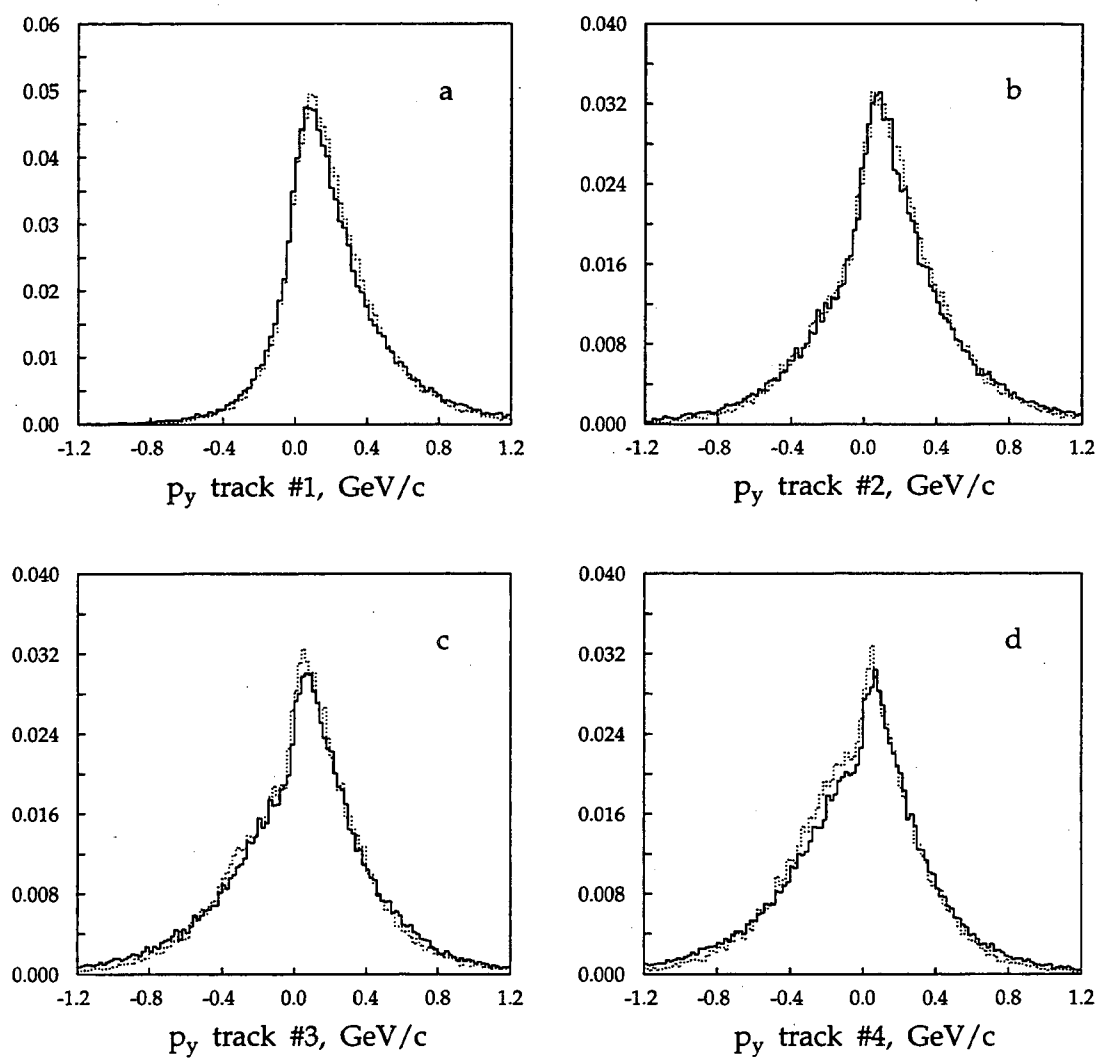


Figure 8.5: Jet-Tagging Neural Network inputs 2 (a), 5 (b), 8 (c), and 11 (d), which are the p_y for tracks 1, 2, 3, and 4, respectively. Solid is bottom, dotted is non-bottom. The area under each curve is 1.

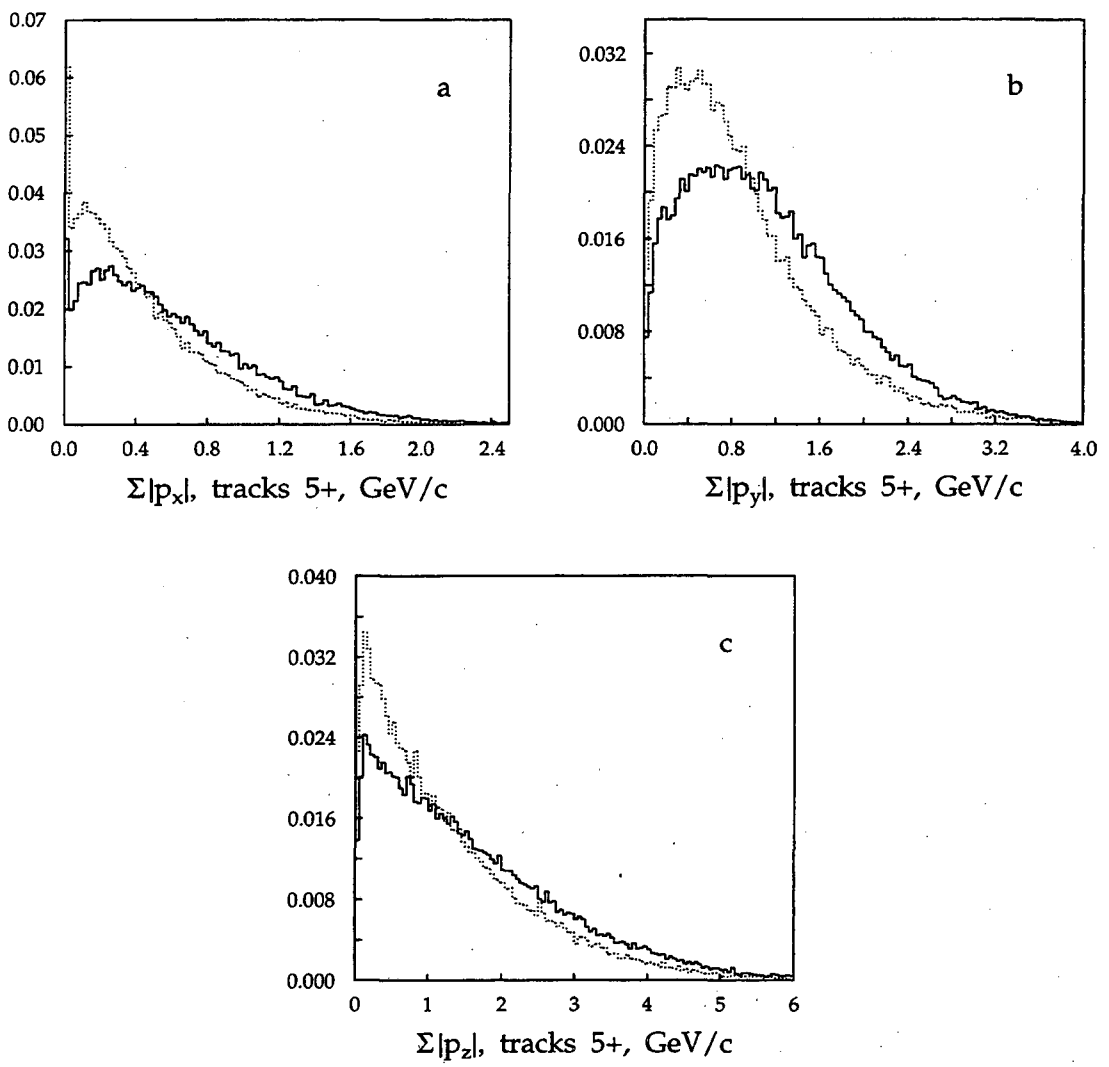


Figure 8.6: Jet-Tagging Neural Network inputs 13 (a), 14 (b), and 15 (c), which are $\sum |p_x|$, $\sum |p_y|$, and $\sum |p_z|$, respectively. The sum is over all tracks but the leading 4. Solid is bottom, dotted is non-bottom. The area under each curve is 1.

purities of the samples above and below the cut. The efficiencies and purities are calculated using the fitted Monte Carlo b and non-b components of the experimental data shown in Figure 8.11. The bin at 1.05 is ignored in the computation of the efficiency and purity: it can be considered a separate sample of jets.

We apply corrections, to the fitted fraction of 0.0747 ± 0.0079, for backgrounds (see Section 7.4.2), for acceptance, and for all of the physics effects discussed in Section 2.7. We find the fraction of bottom jets in $q\bar{q}$ events in QED with zero-mass quarks to be 0.0732 ± 0.0078, which is 2.3σ below the expected fraction of 1/11. We do not find the point where $f_{meas}^{QED} = f_{MC}^{QED}$, as was done in Section 7.6, and we do not compute systematic errors.

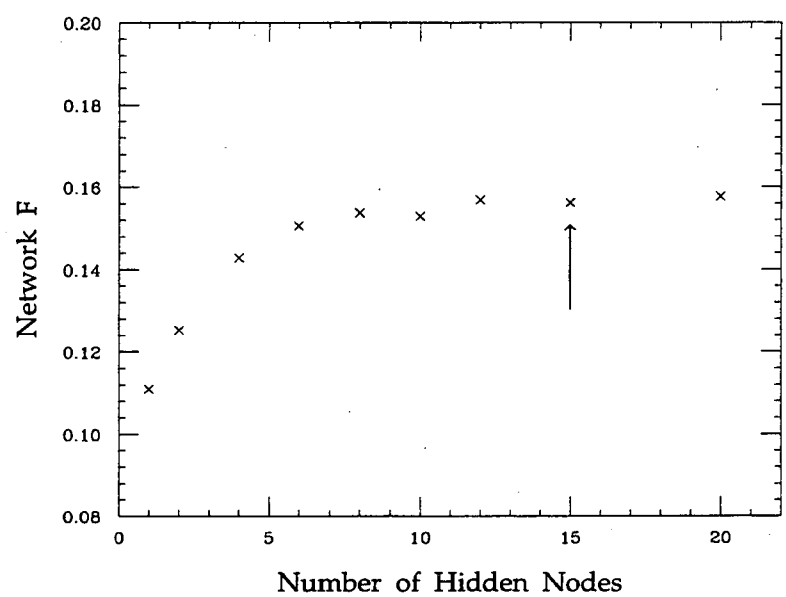


Figure 8.7: F as a function of the number of hidden nodes for the Jet-Tagging Neural Network. The arrow marks the chosen number of hidden nodes (15).

8.4 Techniques for Measuring Charged Hadron Production in Bottom Quark Jets

A conceptually simple approach was used to find the charged hadron momentum spectra for most of the previous analyses described in Section 2.6. The first step in this approach was to use a quark flavor tags to form two samples of tagged jets, each sample having a high purity of one type of quark. The remainder of the entire hadronic data set was used as a third set of jets. Then the charged hadron momentum spectra was found in each of these three samples, and each spectrum corrected for bias and acceptance. Finally, the Monte Carlo-estimated

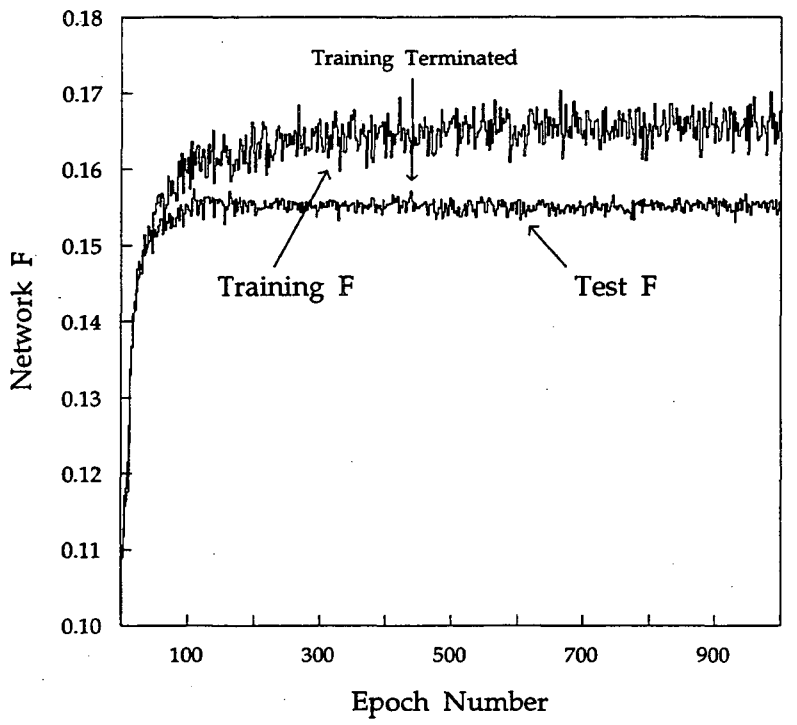


Figure 8.8: F as a function of epoch number for the Jet-Tagging Neural Network.

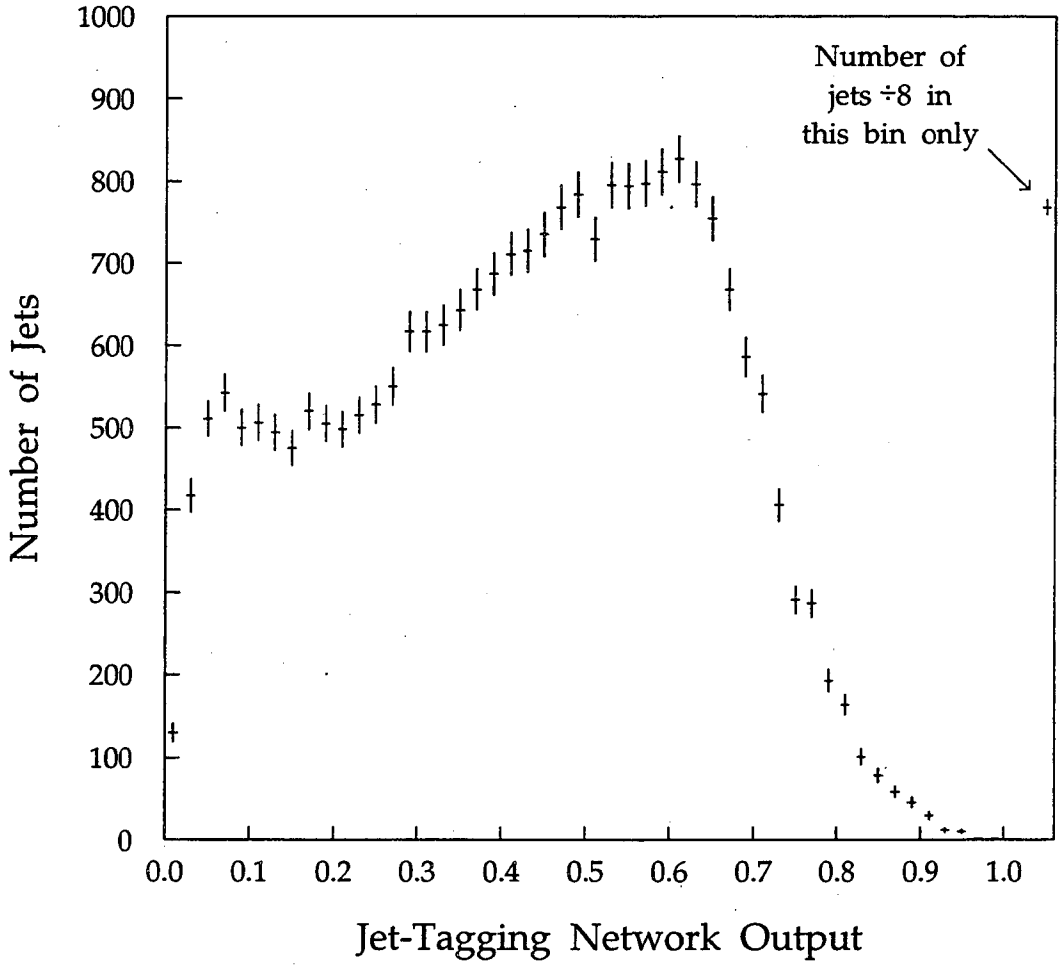


Figure 8.9: Jet-Tagging Neural Network output distribution for Experiments 14-18. The entry in the bin at output 1.05 is the number of jets (÷ 8) with at most 4 tracks.

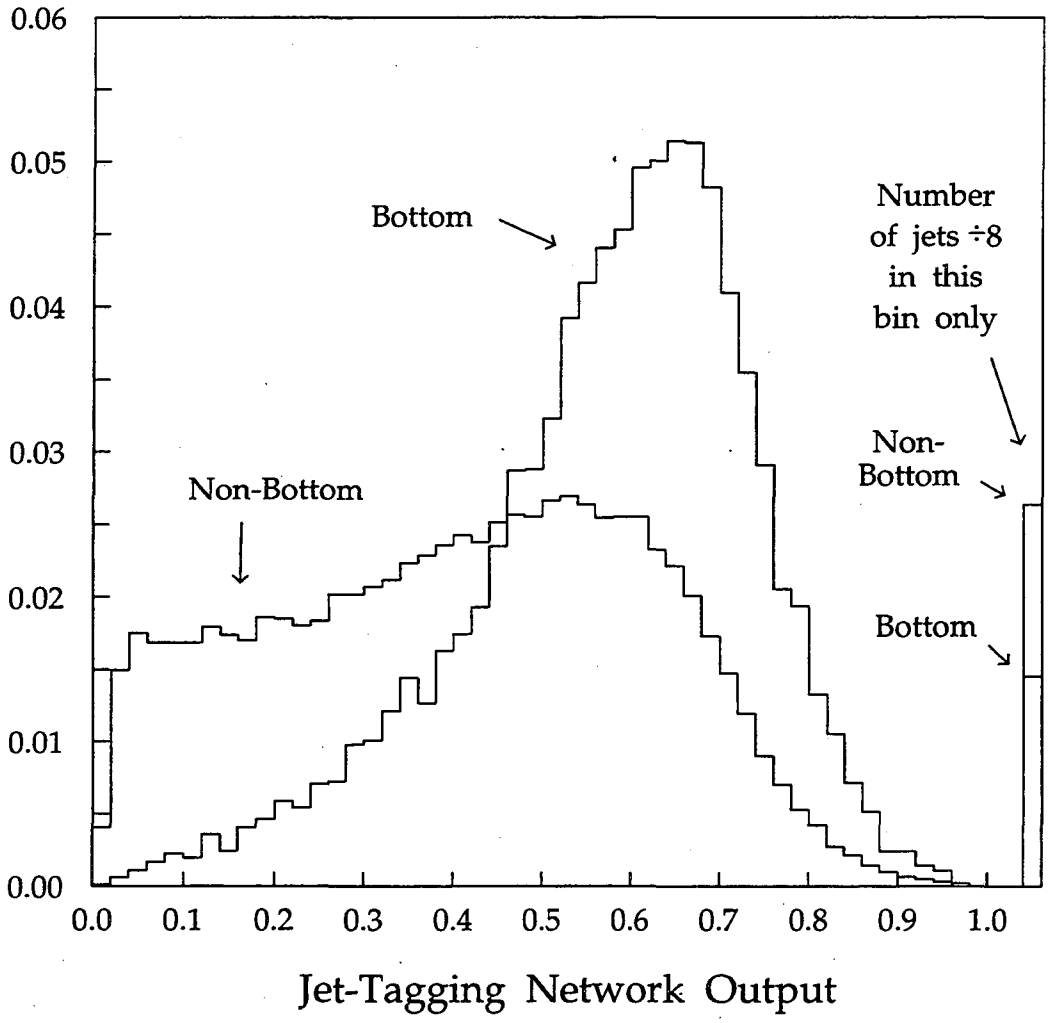


Figure 8.10: Jet-Tagging Neural Network output distributions for Monte Carlo non-bottom and bottom jets. The entries in the bin at output 1.05 are the number of jets (÷ 8) with at most 4 tracks. The area under each curve is 1.

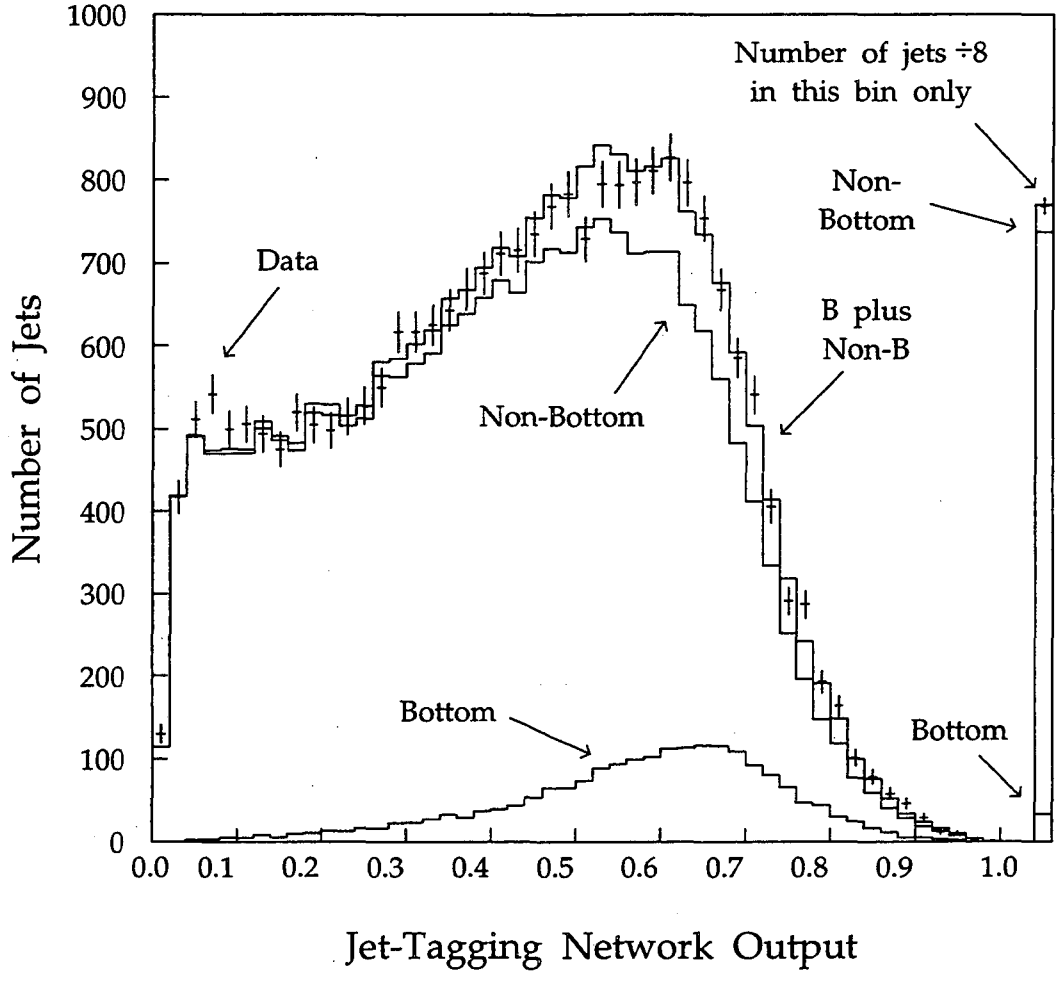


Figure 8.11: The fitted jet-tagging neural network output distributions for Monte Carlo b and non-b jets, their sum, and the output distribution for Experiment 14-18 jets. The entries in the bin at output 1.05 are the number of jets (÷ 8) with at most 4 tracks.

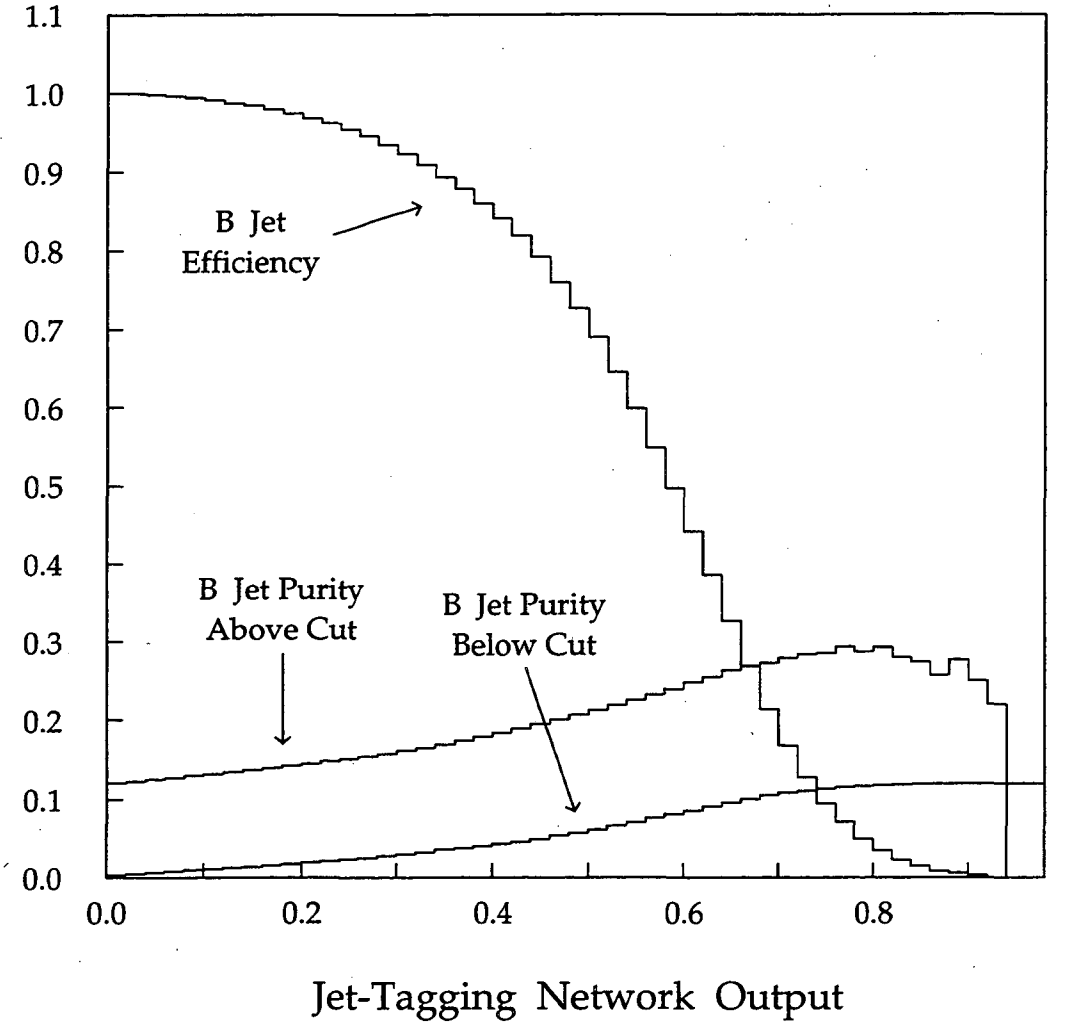


Figure 8.12: Bottom jet efficiency for the sample of jets with network output above the cut on the abscissa. Also shown are the purities of the samples above and below the cut. The bin at output 1.05 is ignored in the computation of the efficiency and purity.

quark content of these three samples was used to extract, from the momentum spectra of these samples, the charged hadron momentum spectra for bottom jets, for charm jets, and for light-quark jets.

This approach is ideal for the tagging methods described in Section 2.6 because each of these tags produce one sample of tagged jets, each sample highly enriched in one quark type. In contrast, a neural network output offers no obvious samples, and Figure 8.12 shows that the sample of jets above any cut on the network output does not have a high bottom jet purity.

We attempted to develop a likelihood function to extract the hadron momentum spectra. There are two motivations for taking this approach. The first motivation is that using two cuts on the neural network output to divide the tagged jets into three samples throws away information present in the distribution of network outputs[1], whereas a likelihood fit does not throw away this information. The second motivation is that a likelihood fit makes use of the information in the differences between the hadron momentum spectra in fitting the quark fractions and hadron momentum spectra, whereas extracting the hadron momentum spectra from three samples of tagged jets only uses the quark contents of the three samples.

One disadvantage of using a likelihood function is that it requires a model for track-track correlations in the tagged jet. We made the approximation that there are no track-track correlations, and we excluded positively charged tracks from the hadron momentum spectra to remove short-range[2] charge and flavor correlations (Bose-Einstein correlations are not removed). This assumption and selection were inadequate, resulting in bias of the fitted hadron momentum spectra in a fit to a Monte Carlo sample. The other disadvantage of using a likelihood function is that bias and acceptance corrections can only made on the fitted hadron momentum spectra, not on the input to the fit. In contrast, it is simple to apply bias and acceptance corrections to the hadron momentum spectra for the three samples before the extraction of the hadron momentum spectra for the different types of quark jets.

We do not pursue the alternative approach of using two cuts on the neural network output to divide the tagged jets into three samples, though there is no reason it can not be done.

[1]This can be seen by coarsening the binning in Figure 8.10 so that there are only three bins.
[2]Short-range in rapidity.

8.5 A Test for Bias in Measurements of Charged Hadron Production in Bottom Quark Jets

In this section, we test for bias in techniques to measure charged hadron production in bottom quark jets by determining if the hadron momentum spectra in Monte Carlo for the tagged jet is independent of the value of the jet-tagging neural network output for the tagging jet. Our test of independence works as follows. We divide the range of the neural network output from 0.0 to 1.0 into 10 bins with width 0.1, place the jets with network output 1.05 into an eleventh bin, and compute the hadron momentum spectra in each of these 11 bins separately for the tracks in bottom, charm, and light-quark (up, down, and strange) Monte Carlo tagged jets that pass all the selections described in Section 8.2. For pions, we use 12 bins in momentum with boundaries .006, .015, .030, .045, .060, .075, .090, .12, .15, .2, .3, .5, and .9 in the scaled momentum $x_p = p/p_{\text{beam}}$. For kaons and protons, we use 4 bins in momentum (because fewer of these particles are produced) with boundaries .006, .045, .090, .2, and .9 in x_p.

For this test, we use 49096 light-quark, 35592 charm, and 74800 bottom jets that pass the selections described in Section 8.2 (Approximately 6.3, 6.3, and 46, respectively, times the number of jets in the Experiment 14-18 data set that pass the same selections.). Once we obtain the hadron momentum spectra in each of the 11 network output bins, we scale the spectra so that each network output bin contains the same total number of hadrons. Finally, we compute the confidence level, in each momentum bin, for the scaled cross-section to be the same in all 11 bins in the neural network output. We compute this confidence level separately for each of the three types of quarks and for each hadron.

We show these confidence levels in Figures 8.13 through 8.15, with light-quark and charm-quark jets separate and combined. The confidence levels for the pion spectra are essentially zero in some momentum bins. The confidence levels for the kaon and proton spectra are larger than for the pion spectra because fewer kaons and protons are produced, but in some momentum bins these kaon and proton confidence levels are about 1% or less. The hadron momentum spectra in the tagged jet obviously depend upon the network output of the tagging jet. Unless there is a miraculous cancellation in the contribution of this dependence to the bottom-jet hadron momentum spectra in each momentum bin, these spectra will be biased[3]. The statement that the spectra are biased applies to any technique of measuring charged hadron production in bottom quark jets that uses our neural network to tag the bottom jets. Since we have taken pains to define the neural

[3]That the confidence levels in Figures 8.13d, 8.14d, and 8.15d, with charm and light-quark jets combined, are sometimes far smaller than the confidence levels for charm and light-quark jets treated separately confirms that the hadron momentum spectra should be found separately for charm-quark and light-quark jets.

network inputs so that bias is minimized, any technique using any neural network will be biased.

There are two ways of correcting for this bias. The simplest way is to make bias corrections. The other way is to try to modify the technique so that the bias corrections are smaller (or so that the confidence levels in Figures 8.13 through 8.15 are closer to 1), and then make the necessary bias corrections.

One possible way of modifying the technique to reduce bias is to exclude the jets with network output 1.05 from the hadron momentum spectra. When these jets are excluded, Figures 8.13 through 8.15 show that the confidence levels for the momentum spectra being independent of the network output improve, in some momentum bins, far beyond the amount expected from the decrease in the statistics. The fact that the improvement mostly takes place in the highest momentum bins of the uds-quark histograms (Figures 8.13a, 8.14a, and 8.15a), where improvement always occurs, suggests that the improvement is produced by the long-range correlation due to the common flavor of the leading quark in each jet.

In the next section, we present a systematic investigation of the possible sources of bias in techniques of measuring charged hadron production in bottom quark jets.

8.6 An Investigation of the Sources of Bias

In Section 8.5, we demonstrated that the hadron momentum spectra of the tagged jets depend upon the network output of the tagging jet. In this section, we present our investigation of the causes of this dependence.

8.6.1 Using Correlations Between Jets in Each Event to Measure Bias

We need an easy-to-compute measure of bias for our investigation of the sources of bias. The easiest, albeit indirect, way to determine how much bias there is for one quark flavor is to see if the neural network inputs in the 2 jets in each event are statistically independent, since this does not require that a neural network be trained. The advantage of this approach is that the inputs can be treated separately, and their behavior compared. This method does not provide an absolute measure of bias.

A necessary, but not sufficient, condition for two variables to be statistically independent is for the correlation to be statistically consistent with being zero. Therefore, we compute the correlation coefficient [120] of the same neural network input in the two jets. In the statistical literature, there are more sophisticated measures of independence than the correlation coefficient, but it is not necessary to use them here.

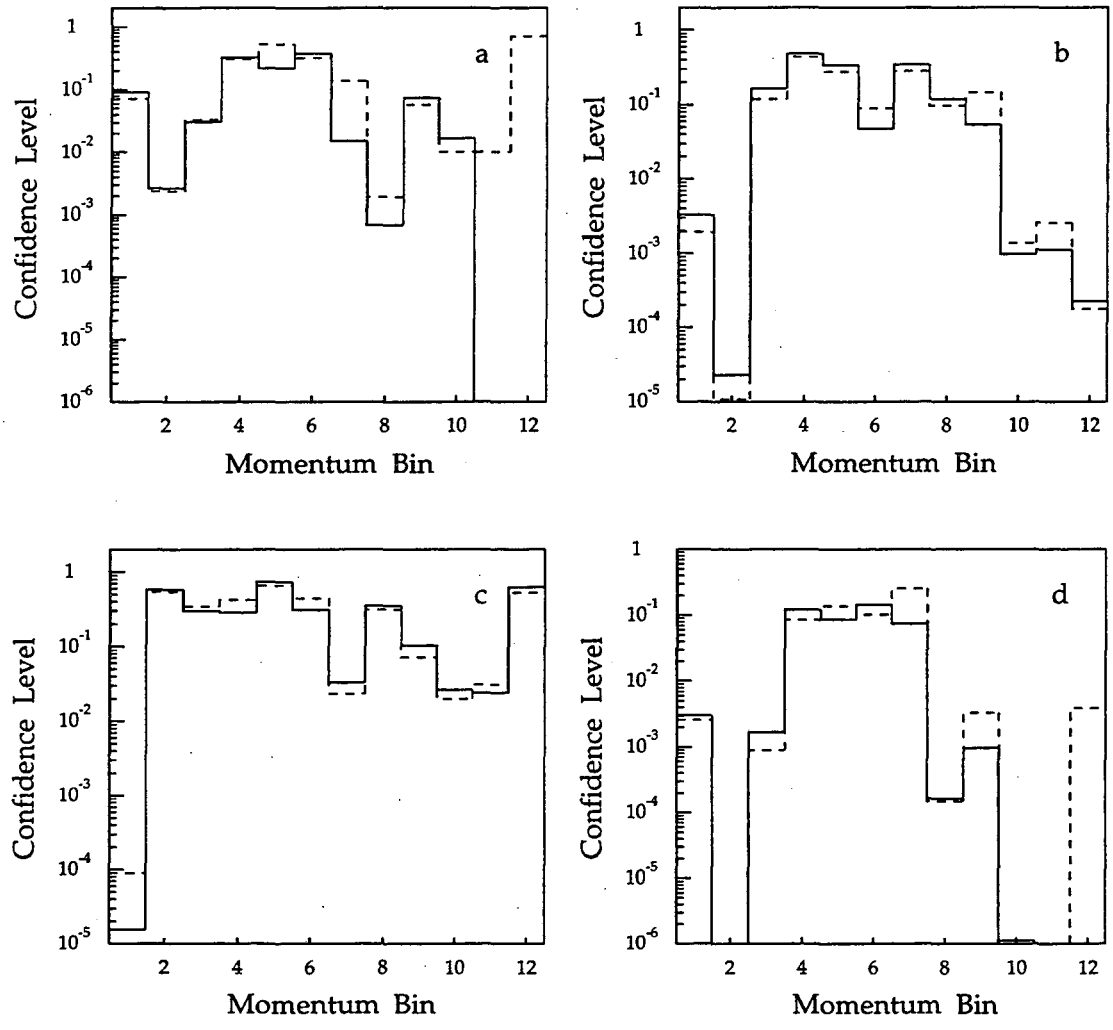


Figure 8.13: The bin-by-bin confidence levels for the scaled pion cross-section to be independent of the network output for Monte Carlo uds-jets (a), charm jets (b), bottom jets (c), and udsc-jets (d). The empty-looking bins really have entries below 10^{-6}. The solid histogram includes jets with network output 1.05, the dashed histogram excludes jets with network output 1.05.

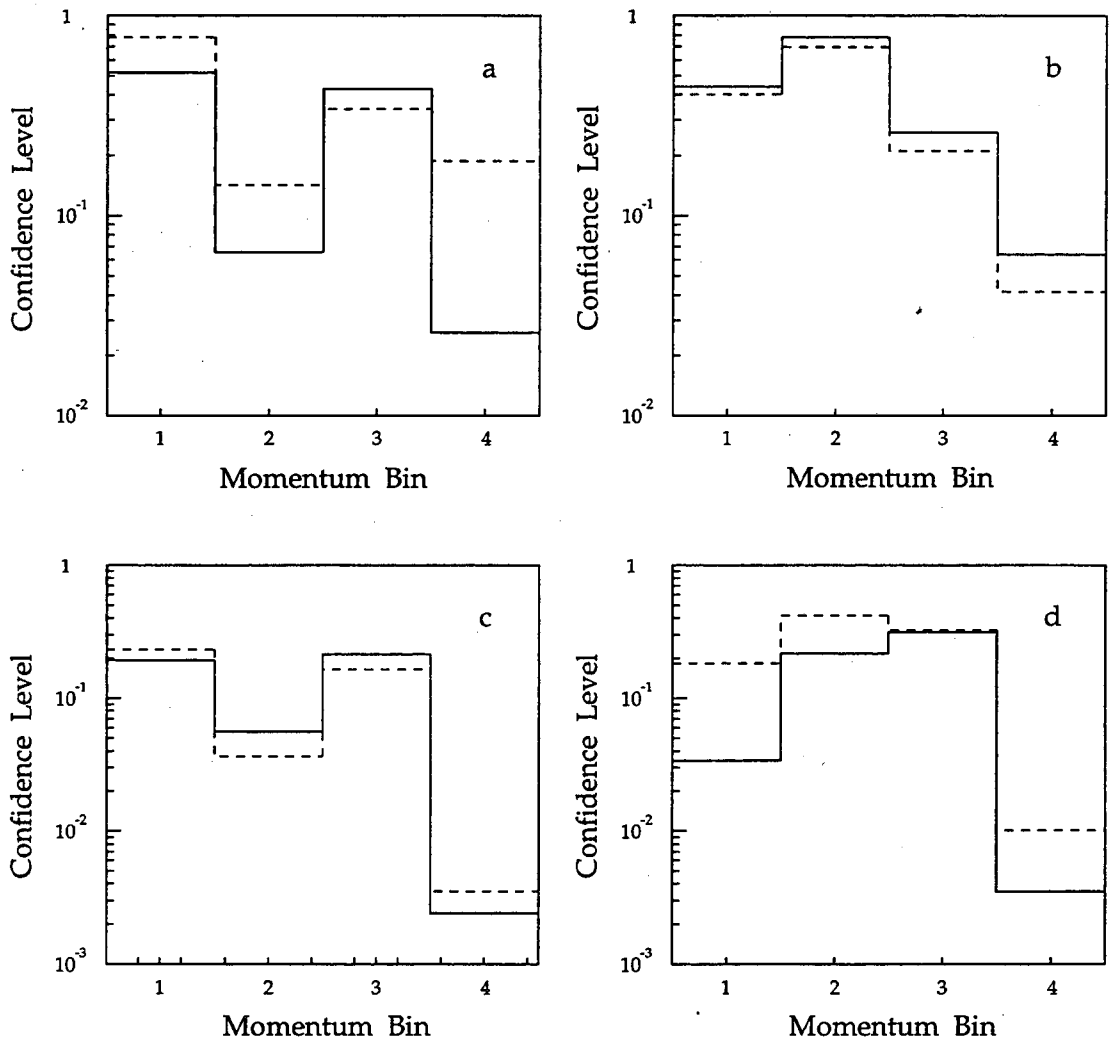


Figure 8.14: The bin-by-bin confidence levels for the scaled kaon cross-section to be independent of the network output for Monte Carlo uds-jets (a), charm jets (b), bottom jets (c), and udsc-jets (d). The solid histogram includes jets with network output 1.05, the dashed histogram excludes jets with network output 1.05.

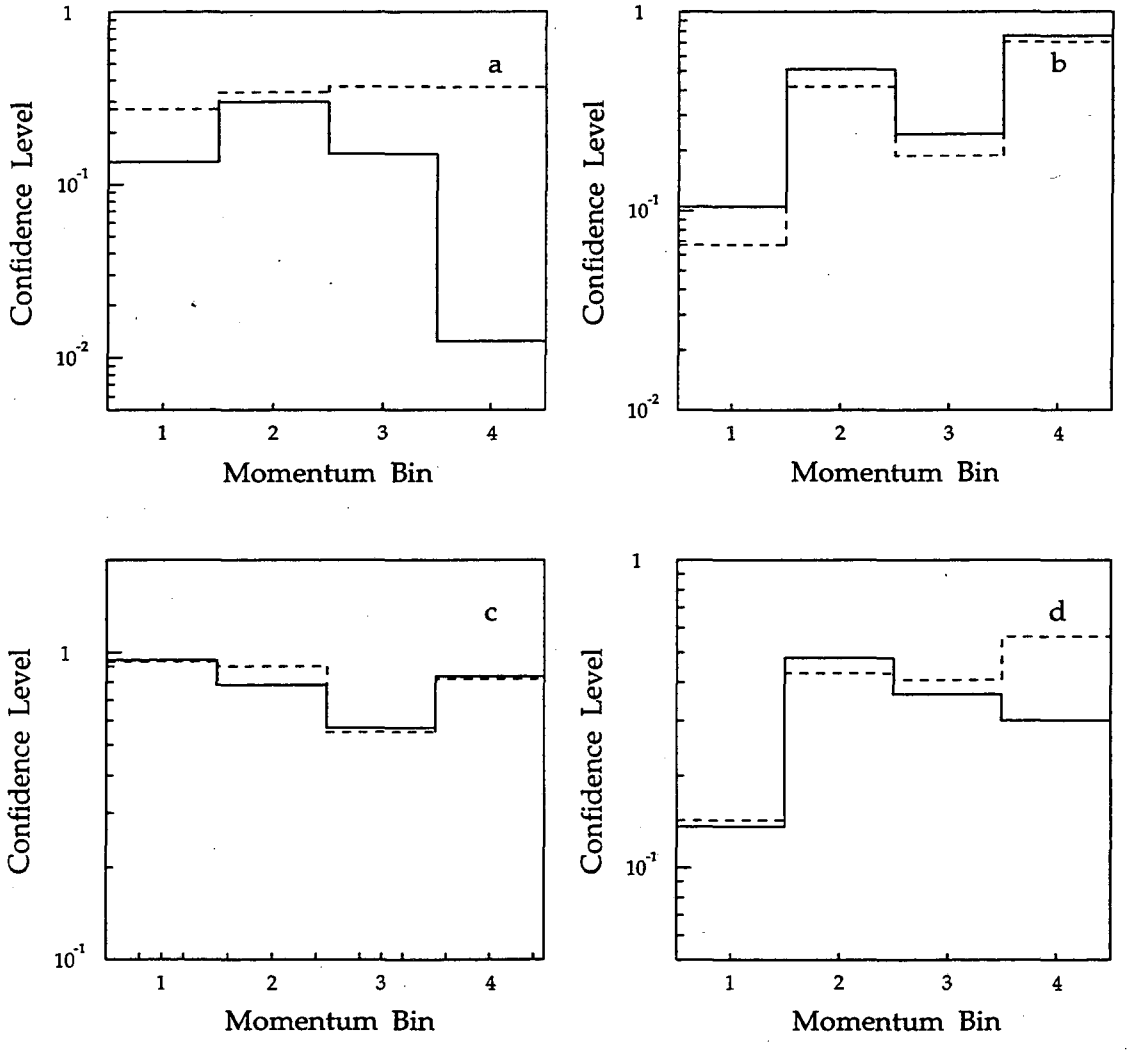


Figure 8.15: The bin-by-bin confidence levels for the scaled proton cross-section to be independent of the network output for Monte Carlo uds-jets (a), charm jets (b), bottom jets (c), and udsc-jets (d). The solid histogram includes jets with network output 1.05, the dashed histogram excludes jets with network output 1.05.

The common quark flavor of the two jets, in combination with the differing neural network input distributions for different quark flavors, will introduce a correlation. This is illustrated in Figure 8.16, where three circular (and therefore independent) distributions in y as a function of x are superposed; the sum of the three distributions has a positive correlation in y as a function of x, since the average of y is a function of x. We are only interested in the correlation due to information other than quark flavor crossing the boundary between the two jets, so we compute separately the correlations for uds, charm, and bottom.

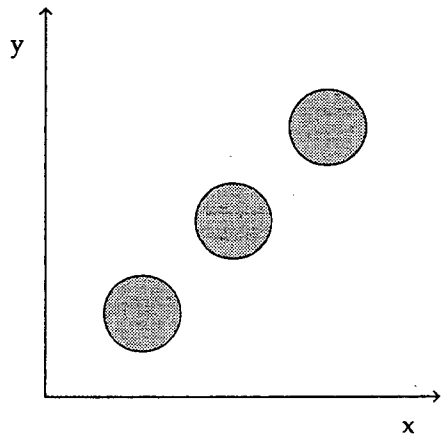


Figure 8.16: Each of the three independent distributions (circles) is independent, but the sum of the distributions is not independent in y as a function of x.

We now have a profusion of correlation coefficients ρ. For the p_x inputs, we have 3 flavors $\times$ 5 inputs = 15 correlation coefficients. We have the same number of coefficients for p_y, p_z, and y. We deal with this by combining this information into three numbers for p_x, three for p_y, three for p_z, and three for y:

1. We average ρ over the 3 quark flavors and the 4 leading tracks (12 terms in the average). We also compute the standard deviation of this average.

2. We compute a χ^2 that the same 12 ρ's (12 degrees of freedom) are all consistent with being zero: $\chi^2 = \sum(\frac{\rho}{\sigma(\rho)})^2$. We also compute the corresponding confidence level.

3. We need some function of only the ρ's to quantify their spread, so we compute $\sum|\rho|$ for the same 12 ρ's.

We use these three quantities to measure how much correlation there is. All of these quantities are computed assuming that the correlations for different tracks in the same jet are uncorrelated with each other. This assumption is not necessarily true, it is certainly not true for y when we order the tracks in y.

We order the tracks in y, therefore the χ^2 C.L. for y is unreliable for deciding whether a correlation exists in y, and we do not consider y correlated unless the χ^2 C.L. is less than 0.5% and unless $(\sum|\rho|)/\sigma(\sum\rho)$ and $\bar{\rho}/\sigma(\bar{\rho})$ are both larger than roughly 3. For p_x, p_y, and p_z, we use the same criteria in order to be conservative.

These criteria are admittedly somewhat arbitrary, but this is not important, since many effects clearly produce large correlations.

Our first task in investigating the sources of correlations is to enumerate all possible sources of correlations, which we do in the next section. After that, we remove all of these correlations from the Monte Carlo and verify that there is no remaining correlation. We then evaluate the correlation that results when each potential source of correlations is present alone (with the other potential sources not present) in the Monte Carlo.

8.6.2 All Possible Sources of Correlations

1. Initial state radiation (ISR).

2. Gluon radiation. All events consist of back-to-back $q\bar{q}$ pairs in the absence of ISR and gluon radiation.

3. Fragmentation/4-momentum-conservation. In order to shut off correlations due to fragmentation, we use independent fragmentation in place of the usual string fragmentation. There are numerous 4-momentum-conservation schemes for independent fragmentation, and we have purposely chosen none of them, so that energy-momentum conservation does not induce correlations. As a result, the total 4-momentum is not conserved, and the total event energy has a mean of 29.0 GeV and a sigma of about 0.6 GeV.

4. All detector effects can introduce correlations, since they are dip-angle dependent and the two jets have a similar dip. Therefore, we have turned off all processes that occur in the material of the TPC/2γ experiment: $\mathrm{d}E/\mathrm{d}x$ energy loss, bremsstrahlung, pair conversion, and nuclear interactions. Some tracks enter the TPC volume more than once, since their paths curl up in the TPC magnetic field, so we ignore all but the first passage of these tracks through the TPC volume.

5. Using a measured quantity, such as the thrust axis, to define a plane for dividing events in half will introduce correlations, since the finding of any such axis uses information from the entire event. Instead, we use the *quark axis*, which is the quark's direction. We also need a direction perpendicular to the quark axis: we pick a perpendicular direction randomly.

6. We remove all particles that go into the half of the event opposite to the parton it originated from. Only in independent fragmentation can one say unambiguously which parton a hadron originated from.

7. It is common to label as jet number 1 the jet with the largest value of some quantity, such as $|\sum p_z|$, the z-direction being the direction parallel to the quark axis. The other jet is jet number 2, and so the correlation of jet 1 with jet 2 is evaluated. This ordering introduces a correlation. In an actual measurement, we would randomly assign the label jet 1. If we use the quark axis, jet 1 is the quark jet and jet 2 is the antiquark jet.

8. All event selections are capable of creating correlations:

   (a) The Good Hadronic Event Selection.

   (b) The τ-2γ event selection.

   (c) The selection of only 2-jet events.

   (d) The cut on the angle between the axes for the two jets in each event.

   (e) All cuts on particle momentum, dip, etc.

   (f) All cuts on event axis dip angle.

8.6.3 Which Sources of Correlation Contribute

In our investigation of the sources of correlation, 120k Monte Carlo events are used (except as stated), whereas the good hadron DSTs for experiments 14-18 contain about 39k events.

8.6.3.1 The Baseline for this Investigation

To start with, we need to compute the measures of correlation in the neural network inputs, with all possible sources of correlations removed, as a baseline against which to compare all variations in the Monte Carlo and in the computing of the network inputs. We do this in the first row of Table 8.1.

Row number two of Table 8.1 shows that, at the level of statistical significance being explored here, introducing backwards-moving tracks does not produce correlations. Removing backwards-moving particles is clearly impossible to do in experimental data, and it is sometimes unwise to do so, as described in Sections 8.6.3.4 and 8.6.3.5. Therefore, in evaluating all correlations, we use backwards-moving particles, and we use the second row of Table 8.1 as the baseline against which we compare all variations of the Monte Carlo and in the computing of the network inputs. We call it "Baseline 1". We have also evaluated (but do not show) the other

variation	quantity	p_x	p_y	p_z	y
without	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.073, .023	.053, .024	.059, .024	.092, .023
backwards	χ^2, C.L.	15.7, 0.21	10.9, 0.54	13.8, 0.31	24.0, 0.020
particles	$\bar{\rho}$	.0003	−.0032	.0004	.0010
	$\sigma(\bar{\rho})$	.0016	.0016	.0016	.0016
with	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
backwards	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
particles	$\bar{\rho}$	.0002	−.0027	.0014	.0047
(Baseline 1)	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015

Table 8.1: The baseline for the investigation of the sources of correlation.

correlations excluding backwards-moving particles, the change in the correlations is never significant.

We go on evaluate the correlation that results when each potential source of correlations is present alone (with the other potential sources not present) in the Monte Carlo. We arrange the sources of correlations in groups. The baseline will be the first row in every comparison table below, and in all tables all conditions that apply in the first row of this table will also apply, except as otherwise stated.

8.6.3.2 The Contribution of Fragmentation Physics

Any realistic model of fragmentation has information flowing from one jet to the other, whether the model is independent fragmentation with whole-event energy-momentum conservation, string fragmentation, etc. (Table 8.2). The E-$\vec{p}$ conservation scheme used for the independent fragmentation in this table is where the momentum imbalance is shared among the particles according to the particles' energy (MST(6) = 1). This table shows that fragmentation induces measurable correlations in p_z and y.

8.6.3.3 The Contribution of Radiation

Initial state photon radiation and final state gluon radiation also cause correlations.

Here, the usual definition for quark axis is problematic, since the initial quark and antiquark are not back-to-back except by accident. Any definition of the quark axis will carry information about the direction and possibly also momentum of the quark jet into the anti-quark jet, and vice versa. There are two simple schemes for defining the quark axis.

fragmentation	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(indep. frag.	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
without E-$\vec{p}$	$\bar{\rho}$	.0002	−.0027	.0014	.0047
conservation)	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
independent	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.110, .023	.123, .023	.172, .024	.245, .023
fragmentation	χ^2, C.L.	124, 0	143, 0	88, 0	206, 0
with E-$\vec{p}$	$\bar{\rho}$	.0067	−.0096	.0132	.0197
conservation	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
string	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.068, .023	.085, .023	.099, .024	.243, .023
fragmentation	χ^2, C.L.	15.8, 0.20	25.8, 0.012	61, 0	198, 0
	$\bar{\rho}$	−.0042	.0035	.0101	.0192
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015

Table 8.2: Correlations from fragmentation physics.

The one we use here defines the quark axis to be the normalized vector in the same direction as the quark momentum minus the anti-quark momentum; this definition explicitly carries momentum information.

The other scheme defines the quark axis to be the normalized vector in the same direction as the unit vector in the quark direction minus the unit vector in the anti-quark direction. This definition does not explicitly carry momentum information, but the directions of these partons are related to their momenta, and thus this definition indirectly carries momentum information. This second definition gives correlations that are slightly larger than the scheme that we show here (Table 8.3).

In an attempt to eliminate the correlations in Table 8.3, we "iterate" the coordinate axes in Table 8.4. "Iterating" the axes (i.e. using the iterated coordinate system) gets rid of the correlations in p_x, and reduces somewhat the correlations in p_y, p_z, and y.

In both Table 8.3 and Table 8.4, the correlations with ISR and gluon radiation together (row 4) are definitely not the sum of the correlations for ISR alone (row 2) and for gluon radiation alone (row 3).

8.6.3.4 The Contribution of Event Shape/Direction Measurements

So far, we have used the perpendicular to the quark axis to divide the event in half. Any method of dividing the event in half in a real measurement uses information from the entire event, and necessarily introduces correlations. Also, when we do

radiation	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(no	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
radiation)	$\bar{\rho}$	.0002	−.0027	.0014	.0047
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
ISR	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.046, .025	.456, .030	.395, .024	4.731, .025
	χ^2, C.L.	6.4, 0.90	997, 0	898, 0	119901, 0
	$\bar{\rho}$	−.0007	.0551	.0389	.4679
	$\sigma(\bar{\rho})$	.0016	.0020	.0016	.0014
parton	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.155, .024	.222, .026	.659, .023	3.926, .023
showers	χ^2, C.L.	117, 0	178, 0	2480, 0	77704, 0
	$\bar{\rho}$	−.0117	−.0171	.0581	.3614
	$\sigma(\bar{\rho})$	.0015	.0019	.0015	.0013
ISR and	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.151, .025	.414, .026	.620, .024	4.776, .022
parton	χ^2, C.L.	123, 0	779, 0	1887, 0	121494, 0
showers	$\bar{\rho}$	−.0138	.0429	.0561	.4295
	$\sigma(\bar{\rho})$	.0016	.0016	.0015	.0013

Table 8.3: Correlations due to ISR and gluon radiation.

radiation	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.046, .023	.060, .023	.062, .023	.072, .023
(no	χ^2, C.L.	6.7, 0.88	10.9, 0.54	12.7, 0.39	13.3, 0.35
radiation)	$\bar{\rho}$	−.0012	.0010	.0025	.0043
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
ISR	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.060, .024	.120, .019	.472, .023	1.298, .022
	χ^2, C.L.	11.9, 0.45	804, 0	1219, 0	6343, 0
	$\bar{\rho}$	.0014	−.0192	.0087	.1003
	$\sigma(\bar{\rho})$	.0015	.0009	.0015	.0013
parton	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.051, .023	.154, .023	.420, .023	1.861, .023
showers	χ^2, C.L.	14.2, 0.29	129, 0	759, 0	11373, 0
	$\bar{\rho}$	−.0001	.0134	.0360	.1523
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0014
ISR and	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.044, .024	.163, .021	.372, .023	1.483, .023
parton	χ^2, C.L.	6.8 , 0.87	617, 0	568, 0	5960, 0
showers	$\bar{\rho}$	−.0007	−.0130	.0221	.1000
	$\sigma(\bar{\rho})$	.0015	.0011	.0014	.0014

Table 8.4: Correlations due to ISR and gluon radiation, axes "iterated".

not "iterate", we use event the axis to define the coordinate system with respect to which we compute the neural network inputs, and this also introduces correlations (Table 8.5). This is one occasion where removing backwards moving particles is unwise, since doing so would alter the thrust axis direction.

Using a thrust axis that is computed from all of the final-state particles generated by the Monte Carlo definitely induces correlations in p_x and p_y that can be erased by iterating the axes. Correlations from using a thrust axis that is computed from the tracks simulated in TPCLUND as recorded by the TPC are much larger, and they can not be entirely erased in p_y and y by iterating.

One must be careful that in using an algorithm that finds the thrust axis, the jets in an event are not ordered by any kinematic criteria, because this ordering will induce correlations.

8.6.3.5 The Contribution of Detector Acceptance

Detector acceptance results in loss of tracks out the ends of the detector. This will cause the observed characteristics of a jet to vary with the dip of the jet, resulting in correlations (Table 8.6).

In Table 8.6, the quark axis is always used. This is another occasion where removing backwards moving particles is unwise, since the acceptance for backwards tracks is not the same as it is for forwards tracks. We have turned off the simulation of interactions in the TPC.

In Table 8.6, the "standard dip and p cuts" are:

1. Track dip $|\lambda| < 60°$.
2. Dip of the event sphericity axis $|\lambda| < 45°$.
3. For tracks in the track block, the momentum extrapolated to the event vertex must be at least 120 MeV/c, and the momentum in the TPC must be at least 100 MeV/c. For particles in the Monte Carlo block, we require $p > 120$ MeV/c.

Re-entries of tracks into the TPC do not induce correlations (row 2). Limiting ourselves to charged particles in computing the network inputs, in conjunction with not cutting backwards tracks, does induce y-correlations (row 3).

Furthermore, rows 4 and 6 are compatible with being identical, thus correlations from using tracks from the track block that enter the TPC fiducial volume are reproduced by the charged tracks in the Monte Carlo block that pass the dip and p cuts.

The large y correlation in row 5 probably comes from the edges of the TPC's fiducial volume, where the event acceptance as a function of the dip of the event

variation	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(quark axis)	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
	$\bar{\rho}$	.0002	−.0027	.0014	.0047
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
quark axis,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.046, .023	.060, .023	.062, .023	.072, .023
"iterated"	χ^2, C.L.	6.7, 0.88	10.9, 0.54	12.7, 0.39	13.3, 0.35
	$\bar{\rho}$	−.0012	.0010	.0025	.0043
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
thrust axis	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.157, .024	.198, .023	.047, .023	.092, .023
	χ^2, C.L.	142, 0	442, 0	7.0, 0.85	23.3, 0.025
	$\bar{\rho}$	−.0127	.0164	.0022	.0062
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
thrust axis,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.050, .024	.075, .023	.070, .024	.076, .023
"iterated"	χ^2, C.L.	7.0, 0.86	13.5, 0.33	12.8, 0.38	14.4, 0.27
	$\bar{\rho}$	−.0008	.0012	.0026	.0045
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
thrust axis,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.480, .031	.284, .030	.107, .031	.332, .031
e-μ-π-K-p only	χ^2, C.L.	285, 0	121, 0	14.6, 0.27	200, 0
	$\bar{\rho}$	.0390	−.0135	.0018	.0300
	$\sigma(\bar{\rho})$	.0025	.0024	.0024	.0024
thrust axis,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.074, .031	.065, .031	.119, .031	.164, .031
e-μ-π-K-p only,	χ^2, C.L.	9.8, 0.64	6.8, 0.87	20.6, 0.057	50, 1.4E-6
"iterated"	$\bar{\rho}$	.0003	.0023	.0060	.0141
	$\sigma(\bar{\rho})$	.0024	.0024	.0024	.0024
thrust axis,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	2.752, .034	2.404, .031	.850, .024	5.785, .025
calculated	χ^2, C.L.	8615, 0	8223, 0	4740, 0	168384, 0
from track	$\bar{\rho}$	−.2212	.1967	.0884	.5672
block	$\sigma(\bar{\rho})$	.0026	.0024	.0016	.0014
thrust axis,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.026, .024	.143, .025	.067, .024	.398, .026
from track	χ^2, C.L.	2.5, 0.998	58, 0	16.5, 0.17	337, 0
block,	$\bar{\rho}$	−.0011	.0121	.0026	.0301
"iterated"	$\sigma(\bar{\rho})$	.0016	.0018	.0016	.0017

Table 8.5: Correlations from event shape/direction measurements.

variation	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(Monte Carlo	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
block, all	$\bar{\rho}$	.0002	−.0027	.0014	.0047
particles)	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
MC block, all	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.066, .023	.049, .024	.065, .023	.083, .023
particles plus	χ^2, C.L.	14.9, 0.25	11.9, 0.45	16.4, 0.17	15.5, 0.21
tracks that	$\bar{\rho}$	.0002	−.0026	.0013	.0046
re-enter TPC	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
Monte Carlo	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.052, .030	.109, .031	.093, .031	.143, .031
block,	χ^2, C.L.	4.7, 0.967	18.3, 0.11	11.0, 0.53	40, 0.00006
e-μ-π-K-p only	$\bar{\rho}$	.0009	.0029	.0044	.0126
	$\sigma(\bar{\rho})$	.0024	.0024	.0024	.0024
MC block,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.096, .039	.128, .040	.134, .041	.170, .040
e-μ-π-K-p only,	χ^2, C.L.	10.1, 0.61	12.8, 0.39	14.3, 0.28	34.8, 0.0005
standard dip	$\bar{\rho}$	−.0017	.0013	.0083	.0111
and p cuts	$\sigma(\bar{\rho})$	.0031	.0031	.0031	.0031
track block	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.142, .034	.087, .033	.162, .032	.776, .037
	χ^2, C.L.	26.2, 0.010	10.0, 0.61	38, 0.00014	470, 0
	$\bar{\rho}$	−.0106	.0007	.0065	.0605
	$\sigma(\bar{\rho})$	.0026	.0023	.0019	.0030
track block,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.090, .038	.138, .040	.087, .036	.192, .042
standard dip	χ^2, C.L.	12.7, 0.39	17.3, 0.14	8.8, 0.72	33, 0.0010
and p cuts	$\bar{\rho}$	−.0012	−.0018	−.0003	.0118
	$\sigma(\bar{\rho})$	.0027	.0029	.0016	.0034

Table 8.6: Correlations due to detector acceptance.

sphericity axis drops off quickly, the track acceptance in p is a function of dip and p, and the track acceptance in dip also is a function of dip and p.

8.6.3.6 The Contribution of Event Selections

In any analysis, various selections must be made to eliminate backgrounds to the desired process $q\bar{q} \rightarrow hadrons$, and to ensure that the bulk of the event enters the fiducial volume of the detector. Any cut that is made on a quantity defined by the entire event, such as charged energy, will introduce correlations (Table 8.7).

One might consider replacing a cut on an event-defined quantity with a cut on each of the event's two jets, for example on the minimum of the total charged energies of the two jets, but the process of dividing the event in half might introduce correlations through the cut, though the correlations might be drastically reduced by using this alternative.

However, altering the selections is not necessary. The χ^2 C.L. indicates that the 2-jet event selection and the jet-jet angle cut induce small correlations. The τ-2γ event selection numbers look interesting, but meet none of the criteria we use for deciding that there are significant correlations. With all selections combined, there is no evidence for correlations.

8.6.3.7 The Contribution of Interactions in Detector Material

Interactions with the material of the TPC/2γ detector are dip dependent. Tracks will be ranged out in the material in front of the TPC, multiple scattering will alter track direction, energy loss will change track momentum, and conversions, nuclear interactions, conversions, and bremsstrahlung (in conjunction with conversions) will add tracks to events. This will cause the observed characteristics of a jet to vary with the dip of the jet, causing the correlations in Table 8.8.

Conversions, nuclear interactions, and bremsstrahlung each satisfy 2 of the 3 criteria for deciding there are significant correlations in y, so we consider these to be borderline cases. $\mathrm{d}E/\mathrm{d}x$ energy loss definitely causes correlations in y, as does the combination of all interactions.

8.6.4 The Relative Importance of the Sources of Correlation

We now compare the size of correlations. The numbers in these tables indicate that the following effects are the sources of correlations, ranked in $\sum |\rho|$ from largest to smallest:

1. ISR and gluon radiation.

selections	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(no	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
selections)	$\bar{\rho}$	.0002	−.0027	.0014	.0047
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
The	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.090, .028	.057, .029	.097, .028	.091,.028
τ-2γ	χ^2, C.L.	17.9, 0.12	6.8, 0.87	27.0, 0.0077	16.2, 0.18
event	$\bar{\rho}$	.0029	−.0022	−.0058	.0006
selection	$\sigma(\bar{\rho})$	.0021	.0021	.0021	.0021
The Good	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.073, .026	.069, .027	.075, .027	.076, .026
Hadronic	χ^2, C.L.	13.4, 0.34	12.3, 0.42	17.0, 0.15	12.2, 0.43
Event	$\bar{\rho}$	.0015	−.0026	−.0021	.0017
Selection	$\sigma(\bar{\rho})$	.0019	.0019	.0019	.0018
2-jet event	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.052, .025	.052, .026	.093, .026	.086, .025
selection	χ^2, C.L.	9.1, 0.70	9.1, 0.69	41.5, 0.00004	16.5, 0.17
	$\bar{\rho}$	.0021	−.0013	−.0049	−.0001
	$\sigma(\bar{\rho})$	.0018	.0018	.0018	.0018
"iterated"	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.088, .044	.114, .041	.093, .041	.208, .040
axes, cut	χ^2, C.L.	6.7, 0.87	10.0, 0.61	8.9, 0.71	40.6, 0.00006
on jet-jet	$\bar{\rho}$	−.0018	−.0033	.0040	.0191
angle	$\sigma(\bar{\rho})$	.0034	.0032	.0032	.0033
cut of |dip|	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.104, .044	.098, .045	.108, .045	.108, .044
$< 60°$ for	χ^2, C.L.	9.4, 0.67	6.0, 0.92	10.2, 0.59	10.7, 0.55
sphericity,	$\bar{\rho}$	.0016	−.0016	.0017	.0056
jet axes	$\sigma(\bar{\rho})$	.0035	.0036	.0036	.0035
all	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.092, .047	.132, .044	.154, .046	.148, .044
selections	χ^2, C.L.	6.8, 0.87	11.1, 0.52	17.6, 0.13	14.6, 0.26
together	$\bar{\rho}$	−.0009	−.0025	.0038	.0037
	$\sigma(\bar{\rho})$	.0037	.0036	.0036	.0036

Table 8.7: Correlations from event selections.

interactions	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(no	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
interactions)	$\bar{\rho}$	.0002	−.0027	.0014	.0047
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
conversions	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.049, .023	.048, .023	.070, .023	.086, .024
	χ^2, C.L.	5.4, 0.94	7.6, 0.82	12.7, 0.39	36.4, 0.00028
	$\bar{\rho}$	−.0006	−.0005	.0016	.0031
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
nuclear	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.056, .023	.059, .023	.073, .023	.084, .024
interactions	χ^2, C.L.	10.8, 0.54	14.9, 0.25	20.1, 0.066	33.7, 0.00074
	$\bar{\rho}$	−.0021	−.0017	.0017	.0029
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
brems-	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.048, .023	.053, .023	.080, .023	.081, .024
strahlung	χ^2, C.L.	7.4, 0.83	9.1, 0.69	17.0, 0.15	29.5, 0.0033
	$\bar{\rho}$	−.0017	−.0010	.0021	.0023
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
multiple	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.052, .023	.069, .023	.041, .023	.056, .023
scattering	χ^2, C.L.	6.6, 0.88	20.4, 0.060	6.4, 0.89	8.6, 0.74
	$\bar{\rho}$	.0000	.0023	.0010	.0020
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
dE/dx	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.081, .023	.064, .023	.053, .023	.079, .023
energy loss	χ^2, C.L.	16.8, 0.16	15.7, 0.21	13.2, 0.36	39.8, 0.00008
	$\bar{\rho}$	−.0008	−.0001	.0025	.0063
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
all	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.054, .020	.067, .020	.063, .020	.085, .020
interactions	χ^2, C.L.	16.8, 0.16	16.6, 0.16	20.0, 0.067	38.7, 0.0001
turned on	$\bar{\rho}$	−.0004	.0022	.0024	.0059
(179k events)	$\sigma(\bar{\rho})$	.0013	.0013	.0013	.0013

Table 8.8: Correlations caused by interactions in the detector material.

2. Use of the thrust axis, when only charged tracks are used.
3. Fragmentation physics.
4. Detector acceptance, the jet-jet angle event cut.
5. The 2-jet event selection.
6. $\mathrm{d}E/\mathrm{d}x$ energy loss, conversions, nuclear interactions, and bremsstrahlung.

Note that the first and third largest sources of correlations are from physics, not from the method used to compute the network inputs, and can not be eliminated. The second-largest source *is* from the method used to compute the network inputs, but this source can not be eliminated since events must be divided in half.

Factors that do not measurably contribute are:

1. The other event selections: the Good Hadronic Event Selection, the τ-2γ selection, and the cut on the |dip| of the event axis.
2. Multiple scattering.
3. Tracks entering the TPC volume more than once.

This summary is not absolute, since what we have done is to have only one potential source of correlations present at a time. Correlations do not add linearly: some contributions no doubt cancel and others reinforce each other. It may be that some contributions that are minor or negligible in this ranking make large contributions when combined with other sources of correlations. Or it may happen that some sources of correlations, when combined, do not have measurable correlations, such as when all event selections are combined.

This can be checked by evaluating correlations with all of our supposedly non-contributing candidate sources of correlations present: track re-entries into the TPC, multiple scattering, and the Good Hadronic Event Selection, the τ-2γ event selection, and the cut on the |dip| of the event axis. We call this combination "Baseline 2". In Table 8.9, we evaluate the correlations in Baseline 2, alone and combined with other interactions and event selections. The axes are iterated.

There are no correlations, either in Baseline 2 alone or when all interactions and all event selections are included. Therefore, we use Baseline 2 plus all interactions and all event selections as a new baseline. We call this "Baseline 3".

We can add on to Baseline 3 the contributing sources of correlations, one at a time, and see if the previous ranking still holds (Table 8.10).

When contributing sources of correlations are combined, one at a time, with all non-contributing potential sources of correlations, their size and ranking change

sources	quantity	p_x	p_y	p_z	y
Baseline 1	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.065, .023	.063, .024	.065, .023	.084, .023
(no sources)	χ^2, C.L.	19.0, 0.088	16.1, 0.19	16.2, 0.18	16.3, 0.18
	$\bar{\rho}$	.0002	−.0027	.0014	.0047
	$\sigma(\bar{\rho})$	.0015	.0015	.0015	.0015
Baseline 2	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.150, .044	.104, .044	.141, .044	.137, .044
	χ^2, C.L.	15.3, 0.23	9.2, 0.69	12.3, 0.42	20.1, 0.065
	$\bar{\rho}$	−.0028	.0016	.0027	.0116
	$\sigma(\bar{\rho})$	.0036	.0035	.0035	.0036
Baseline 2	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.155, .045	.107, .044	.135, .044	.138, .044
plus other	χ^2, C.L.	16.0, 0.19	9.9, 0.63	11.9, 0.45	20.0, 0.068
event	$\bar{\rho}$	−.0029	.0018	.0024	.0117
selections	$\sigma(\bar{\rho})$	.0036	.0035	.0035	.0036
Baseline 2	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.098, .033	.076, .034	.058, .033	.118, .033
plus other	χ^2, C.L.	10.5, 0.57	9.6, 0.65	5.4, 0.94	20.1, 0.065
interactions	$\bar{\rho}$	.0006	.0001	−.0013	.0056
(179k events)	$\sigma(\bar{\rho})$	.0026	.0026	.0026	.0026
Baseline 2 + other	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.094, .033	.078, .034	.061, .033	.115, .033
interactions and	χ^2, C.L.	9.8, 0.63	9.8, 0.63	5.5, 0.94	18.4, 0.10
event selections	$\bar{\rho}$	.0005	.0001	−.0013	.0052
(179k events)	$\sigma(\bar{\rho})$	.0026	.0026	.0026	.0026

Table 8.9: Correlations relative to "Baseline 2".

sources	quantity	p_x	p_y	p_z	y
Baseline 3,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.134, .041	.101, .041	.170, .040	.247, .040
plus	χ^2, C.L.	18.5, 0.10	9.6, 0.65	44, 0.000018	85, 0
detector	$\bar{\rho}$	.0018	.0012	−.0132	−.0213
acceptance	$\sigma(\bar{\rho})$	.0030	.0030	.0030	.0030
Baseline 3,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.116, .047	.085, .046	.171, .045	.129, .046
plus	χ^2, C.L.	8.2, 0.77	5.1, 0.95	29.8, 0.0030	15.1, 0.24
thrust axis	$\bar{\rho}$	.0025	−.0011	−.0050	.0004
	$\sigma(\bar{\rho})$	.0029	.0028	.0028	.0028
Baseline 3,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.130, .039	.149, .038	.171, .039	.136, .039
plus LUND	χ^2, C.L.	13.8, 0.032	20.0, 0.066	35, 0.0004	27.6, 0.0063
fragmen-	$\bar{\rho}$	.0008	−.0016	−.0008	.0097
tation	$\sigma(\bar{\rho})$	.0025	.0025	.0025	.0025
Baseline 3,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.106, .048	.115, .049	.208, .048	.146, .048
plus ISR	χ^2, C.L.	10.8, 0.54	10.2, 0.60	31.8, 0.0015	32.2, 0.0013
	$\bar{\rho}$	−.0004	−.0038	.0014	.0016
	$\sigma(\bar{\rho})$	.0029	.0029	.0029	.0029
Baseline 3,	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.138, .049	.165, .052	.177, .051	.267, .049
plus gluon	χ^2, C.L.	15.6, 0.21	20.4, 0.059	48, 3E-6	69, 0
radiation	$\bar{\rho}$	.0047	.0022	−.0136	.0056
	$\sigma(\bar{\rho})$	.0029	.0029	.0029	.0029

Table 8.10: Correlations relative to "Baseline 3".

drastically. Specifically, the size of the ISR and gluon radiation correlations shrink a lot, the detector acceptance correlation is much bigger, and the thrust axis correlation is larger in p_z and smaller in y. The new ranking, from largest to smallest, is as follows.

1. Gluon radiation.
2. Detector acceptance.
3. ISR.
4. Use of the thrust axis, with all particles, may not induce significant correlations.

The first and third largest sources of correlations are from physics, and can not be eliminated. The second largest source of correlations is from detector acceptance, which is intrinsic to any experiment and can not be eliminated. We use this last ranking as the ranking of the relative importance of the potential sources of correlations.

8.6.5 Comparing Monte Carlo Correlations to Experimental Data Correlations

We can compute the correlations we see in the Monte Carlo we use for simulating the experimental data, with the network inputs computed exactly as we would in a measurement using experimental data. The resulting correlations are listed in Table 8.11.

	quantity	p_x	p_y	p_z	y
uniterated	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.654, .041	1.098, .038	.286, .038	1.560, .039
	χ^2, C.L.	347, 0	1426, 0	98, 0	2994, 0
	$\bar{\rho}$	.0544	−.1020	.0171	.1449
	$\sigma(\bar{\rho})$	.0030	.0028	.0022	.0027
iterated	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.096, .040	0.127, .040	.227, .038	0.792, .040
	χ^2, C.L.	9.9, 0.62	16.8, 0.16	82, 0	723, 0
	$\bar{\rho}$	.0023	.0033	.0136	.0720
	$\sigma(\bar{\rho})$	.0029	.0029	.0022	.0029

Table 8.11: The effect of axis "iterating" on correlations.

There are obvious correlations in all 4 variables with the axes uniterated. Iterating the axes helps reduce correlations, as it usually does, but does not eliminate correlations in p_z and y.

One would like to compare input correlations for experimental data to the correlations for Monte Carlo. This is not possible with the usual methods used, since they look at the correlations for individual flavors. What can be used for comparison are the correlations for all flavors combined, since this is the only way experimental data comes. Therefore, we compute the analogs of the comparison quantities, with all flavors combined (Table 8.12). These correlations can not be compared to the other correlations listed elsewhere, because they are computed differently: these χ^2's are for 4 degrees of freedom.

	quantity	p_x	p_y	p_z	y
uniterated	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.217, .012	.411, .011	.113, .011	.651, .011
Monte Carlo	χ^2, C.L.	335, 0	1363, 0	121, 0	3684, 0
	$\bar{\rho}$	.0543	−.1025	.0239	.1638
	$\sigma(\bar{\rho})$	.0030	.0028	.0026	.0027
uniterated	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.273, .021	.423, .020	.141, .024	.677, .019
Expt 14-18	χ^2, C.L.	171, 0	443, 0	42, 0	1270, 0
	$\bar{\rho}$	.0688	−.1046	.0319	.1704
	$\sigma(\bar{\rho})$	.0053	.0050	.0055	.0049
iterated	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.013, .012	.025, .012	.101, .010	.380, .011
Monte Carlo	χ^2, C.L.	2.3, 0.68	7.1, 0.13	108, 0	1125, 0
	$\bar{\rho}$	.0025	.0036	.0204	.0953
	$\sigma(\bar{\rho})$	.0029	.0029	.0025	.0029
iterated	$\sum\|\rho\|$, $\sigma(\sum\rho)$	.028, .021	.032, .020	.136, .021	.379, .021
Expt 14-18	χ^2, C.L.	3.8, 0.43	3.0, 0.56	42, 0	343, 0
	$\bar{\rho}$	.0071	.0060	.0329	.0947
	$\sigma(\bar{\rho})$	.0052	.0051	.0053	.0052

Table 8.12: Correlations in the Jet-Tagging Network Inputs, for Monte Carlo and for experimental data.

For both iterated and uniterated axes, the correlations in experimental data and Monte Carlo are compatible with being the same, using $\sum|\rho|$, with estimated error $\sigma(\sum\rho)$, and $\bar{\rho}$, with estimated error $\sigma(\bar{\rho})$, to do the comparison. In reading this table, remember that the sizes of the data and Monte Carlo samples are quite different, so the errors on the quantities in this table have different sizes.

We can also test for bias by computing the correlation between the network outputs for the two jets in each event. Networks have been trained with and without iterating, with the usual 15 inputs, one hidden layer with 15 nodes, and 1 output node that is trained to separate b-jets from non-b jets. The resulting correlations, for separated and for combined quark flavors, are listed in Table 8.13. One can see that iterating achieves a statistically significant reduction in the correlations for all quark flavors combined and for uds, c separately. Also, the correlations for b-jets are significantly smaller than for the other flavors.

data set	uniterated ρ	iterated ρ
Expt 14-18	$.180 \pm .010$	$.139 \pm .010$
udscb Monte Carlo	$.195 \pm .005$	$.160 \pm .006$
uds Monte Carlo	$.130 \pm .008$	$.088 \pm .008$
c Monte Carlo	$.151 \pm .009$	$.117 \pm .009$
b Monte Carlo	$.035 \pm .008$	$.028 \pm .009$

Table 8.13: Correlations in the Jet-Tagging Network Output, for Monte Carlo and for experimental data.

Correlations definitely exist, and the correlations for experimental data and for Monte Carlo agree within statistics.

8.7 Summary

The study we present in this chapter of bias in techniques to measure inclusive $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$ production in bottom jets provides a number of results that should be useful for future attempts to tag bottom quark jets using a neural network.

1. We describe a neural network with 15 inputs that we calculate from the 3-momenta of the observed charged hadrons in the tagging jet. We describe a coordinate system, for each jet, for computing the momentum components of the tracks. This coordinate system minimizes correlations between the tagging and tagged jets in each event.

2. We demonstrate that the hadron momentum spectra, in Monte Carlo, for the tracks in tagged jets from different quarks depends upon the value of the network output of the tagging jet. This indicates that bias exists in the measurement of the hadron momentum spectra for tagged bottom jets, where the network output is used to identify bottom jets.

3. We show that the dependence of the tagged jet's hadron momentum spectra upon the value of the tagging jet network output is caused by correlations between the two jets in each event (we only use 2-jet events). These correlations exist between the network outputs for the two jets, and between the same network input in each of the two jets.

4. We show that the Monte Carlo simulation described in Section 5.3.1 successfully reproduces the magnitude of the correlations observed in experimental data for the network inputs and output.

5. We have compiled a list of all possible sources of correlations, and we have carried out a systematic study of the relative importance of these sources of correlation. We show that the largest contributors to the correlations are, in decreasing order of importance,

   (a) Gluon radiation.

   (b) Detector acceptance.

   (c) Initial state photon radiation.

   These three sources of correlation are intrinsic to any measurement and can not be eliminated.

Chapter 9

Conclusions

We have presented a measurement of the bottom event fraction $\sigma(b\bar{b})/\sigma(q\bar{q})$ in the annihilation process $e^+e^- \rightarrow b\bar{b} \rightarrow hadrons$ at $\sqrt{s} = 29$ GeV, corrected for all physics that change this fraction from its value in QED with massless fermions. The fraction we measure is $10.2^{+2.0+2.7}_{-1.8-2.2}\%$ (the first error statistical and the second systematic) $= 10.2^{+3.6}_{-2.8}\%$, which is consistent with the value of 1/11 predicted by the Standard Model. The analysis is based on 66 pb^{-1} of data collected between 1984 and 1986 with the TPC/2γ detector at PEP. To identify bottom events, we use a neural network with inputs that are computed from the 3-momenta of all of the observed charged hadrons in each event. We have shown that the measured fraction depends upon the value of this fraction in Monte Carlo. We obtain the value of the bottom event fraction we report by noting that since the bottom event fraction is a physical observable, the fraction in Monte Carlo must be the same as the expectation value of the fraction measured in experimental data.

We also presented a study of bias in techniques for measuring inclusive $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$ production in the annihilation process $e^+e^- \rightarrow b\bar{b} \rightarrow hadrons$ at $\sqrt{s} = 29$ GeV, using a neural network to identify bottom-quark jets. We described a neural network with 15 inputs, computed from the 3-momenta of all of the observed charged hadrons in each jet, that was designed to minimize correlations between the tagging and tagged jets in each event. We demonstrate that the hadron momentum spectra, in Monte Carlo, for the tracks in the tagged jet depend upon the value of the network output of the tagging jet: an indication of bias. We show that the dependence of the tagged jet's hadron momentum spectra upon the value of the tagging jet network output is caused by correlations between the two jets in each event, and we demonstrate that Monte Carlo successfully reproduces the magnitude of the correlations observed in experimental data in the network inputs and output. Finally, we have carried out a systematic study of the relative importance of all possible sources of correlation, and we show that the largest contributors to the correlations are, in decreasing order of importance,

1. Gluon radiation.
2. Detector acceptance.
3. Initial state photon radiation.

These three sources of correlation are intrinsic to any measurement and can not be eliminated.

Appendix A

Remainder of the Proof of the Event Fraction Likelihood Function Optimization Method

A.1 The Case some but not all $m_{ij} = 0$

In this case, for those values of i for which $m_{ij} = 0$, Equation 7.11 implies

$$u_{ij} = \left(\frac{n_j}{\lambda_j} - 1\right) a_i - 1 \ . \tag{A.1}$$

It is now possible for f_{ij} to be zero, for values of i where $m_{ij} = 0$, at the maximum of $(\ell_E)_j$. In order for f_{ij} to be zero, $u_{ij} < 0$ is necessary. Note that $u_{ij} < 0$ implies Equation 7.13, so this equation always holds whether or not $m_{ij} = 0$.

To find the solution, assume that $f_{ij} = 0$ for those values of i for which $m_{ij} = 0$, use Equation 7.14 to produce an equation for λ_j, solve for λ_j using the method described in Reference [117], and use Equation 7.12 to solve for the other f_{ij}. There are two subcases.

A.1.1 The Subcase $\frac{n_j}{\lambda_j} - 1 < \frac{1}{a_{M'}}$

In this case, the f_{ij} and λ_j we have obtained maximize $(\ell_E)_j$. The proof goes as follows. Equation A.1 and $a_i \geq 0$ imply

$$u_{ij} = \left(\frac{n_j}{\lambda_j} - 1\right) a_i - 1 < \frac{a_i}{a_{M'}} - 1 \leq \frac{a_{M'}}{a_{M'}} - 1 = 0 \tag{A.2}$$

for those values of i for which $m_{ij} = 0$. $f_{ij} = 0$ is thus required. However, Equation 7.12 tells us that $m_{ij} = 0$ implies $f_{ij} = 0$. Therefore, the solution method for the case $m_{ij} \neq 0$ for all i can be applied in this subcase.

A.1.2 The Subcase $\frac{1}{a_{M'}} \leq \frac{n_j}{\lambda_j} - 1$

In this case, Equation 7.13 is violated for $i = M'$. Since we have restricted m_{ij} and f_{ij} so that they do not have opposite signs, this means that $m_{M'j} = 0$ and

$$u_{M'j} = \left(\frac{n_j}{\lambda_j} - 1\right) a_{M'} - 1 > \left(\frac{a_{M'}}{a_{M'}} - 1\right) - 1 = 0 \ . \tag{A.3}$$

Therefore, $(\ell_E)_j$ can be made larger by making $f_{M'j} > 0$. The solution can be obtained by setting

$$\lambda_j = n_j \frac{a_{M'}}{1 + a_{M'}} \tag{A.4}$$

and using Equation 7.12 to solve for f_{ij} for $i \neq M'$. Finally, from the definition $\lambda_j = \sum_i a_i f_{ij}$,

$$f_{M'j} = \frac{\lambda_j - \sum_{i \neq M'} a_i f_{ij}}{a_{M'}} \ . \tag{A.5}$$

This definition of λ_j, when inserted into Equation A.1, implies that $u_{M'j} = 0$. For all other i's with $m_{ij} = 0$, we get the required condition that $f_{ij} = 0$ and u_{ij} is negative, since

$$u_{ij} = \left(\frac{n_j}{\lambda_j} - 1\right) a_i - 1 < \left(\frac{n_j}{\lambda_j} - 1\right) a_{M'} - 1 = u_{M'j} = 0 \ . \tag{A.6}$$

For all other i's with $m_{ij} \neq 0$, the corresponding values of f_{ij} were obtained using Equation 7.12, which is equivalent to u_{ij} in Equation 7.11 set equal to zero. The proof that $f_{M'j}$ is positive is complicated and is left to Reference [117].

A.2 The Case $m_{ij} = 0$ for all i

The solution is that all f_{ij} are zero, except for $i = M'$ where

$$f_{M'j} = \frac{n_j}{1 + a_{M'}} \ . \tag{A.7}$$

This is so because $n_j > 0$ requires that $\lambda_j > 0$, which in turn implies that least one f_{ij} be larger than zero. For these non-zero f_{ij}, u_{ij} is zero, implying

$$\frac{n_j}{\lambda_j} - 1 = \frac{1}{a_i} \ . \tag{A.8}$$

This equation can hold for only one value of i, for the other values of i Equation 7.13 holds. This one value of i is M', since $1/i < 1/M'$ for $i \neq M'$, implying that Equation 7.13 must hold for $i \neq M'$. Then Equation A.8 and $\lambda_j = a_{M'} f_{M'j}$ imply Equation A.7. Of course, if all of the m_{ij} and n_j are zero, then all of the f_{ij} and λ_j are zero, and Equation A.7 still holds.

Appendix B

F for Any Number of Classes

This case is not addressed in References [100] and [101].

Suppose we have K normalized distributions $s_{c,i}$, where $c = 1, 2, \ldots, K$ is the class index. α_c are the corresponding fractions, so $\sum_c \alpha_c = 1$.

Neglecting the Monte Carlo statistical error, our χ^2 is now

$$\chi^2 = N \sum_i \frac{(\sum_c \alpha_c s_{c,i} - d_i)^2}{d_i} \,. \tag{B.1}$$

Rather than explicitly put the constraint $\sum_c \alpha_c = 1$ into the expression for χ^2, eliminate one variable, and ruin the symmetry of χ^2 in the index c, we choose to use Lagrange multipliers as follows.

First, we create a new variable $Z = \chi^2 + \lambda(\sum_c \alpha_c - 1)$. Then we set to zero the derivatives of Z with respect to λ and with respect to the α_c, and solve for the optimized α_c and λ. Then, we create the matrix $\mathbf{M}$ of $\frac{1}{2}$ the 2nd derivatives of Z with respect to α_c and λ, evaluated at the optimized values of α_c and λ. Finally, we invert $\mathbf{M}$. The submatrix of $\mathbf{M}^{-1}$, composed of those elements $(\mathbf{M}^{-1})_{ij}$ such that $\mathbf{M}_{ij} = \frac{\partial^2 Z}{\partial \alpha_{c_1} \partial \alpha_{c_2}}$, is the covariance matrix $\mathbf{V}$ of the α_c [121].

Evaluating $\mathbf{M}$,

$$\mathbf{M}(\alpha_{c_1}, \alpha_{c_2}) = \frac{1}{2} \frac{\partial^2 Z}{\partial \alpha_{c_1} \partial \alpha_{c_2}} = N \sum_i \frac{s_{c_1,i} s_{c_2,i}}{d_i} \tag{B.2}$$

$$\mathbf{M}(\alpha_c, \lambda) = \frac{1}{2} \frac{\partial^2 Z}{\partial \alpha_c \partial \lambda} = \frac{1}{2} \tag{B.3}$$

$$\mathbf{M}(\lambda, \lambda) = \frac{1}{2} \frac{\partial^2 Z}{\partial \lambda^2} = 0 \,. \tag{B.4}$$

As one can see, $\mathbf{M}$ is independent of the α_c and λ. The algebraic inversion of $\mathbf{M}$ in the general case is very complicated except for $K = 2$, where we get

$\mathbf{M}(\alpha,\alpha) = \mathbf{M}(1-\alpha, 1-\alpha) = \sigma^2(\alpha)$, and $\sigma(\alpha)$ takes on the value in Equation 6.10, as expected. It turns out that we only need to algebraically invert $\mathbf{M}$ for the case that follows.

B.1 K Event Classes with Complete Separation

For the case where there is only one class of events in each bin,

$$\mathbf{M}(\alpha_{c_1}, \alpha_{c_2}) = N \sum_i \frac{s_{c_1,i} s_{c_2,i}}{d_i} = N \delta_{c_1,c_2} \sum_i \frac{s_{c_1,i}^2}{d_i} . \tag{B.5}$$

Let $1/Q_c = \sum_i \frac{s_{c,i}^2}{d_i}$, then

$$\mathbf{M}(\alpha_{c_1}, \alpha_{c_2}) = N \frac{\delta_{c_1,c_2}}{Q_{c_1}} . \tag{B.6}$$

Most of the non-diagonal elements are zero, therefore it is not hard to invert $\mathbf{M}$ for any K. Doing so yields the covariance matrix

$$\mathbf{V}_{c,c} = \frac{1}{N} Q_c (1 - Q_c) \tag{B.7}$$

$$\mathbf{V}_{c_1,c_2} = -\frac{1}{N} Q_{c_1} Q_{c_2} \quad \text{for} \quad c_1 \neq c_2 \tag{B.8}$$

$$\sigma_c = \sqrt{\mathbf{V}_{c,c}} = \sqrt{\frac{Q_c(1-Q_c)}{N}} \tag{B.9}$$

$$\rho_{c_1,c_2} = \frac{\mathbf{V}_{c_1,c_2}}{\sigma_{c_1}\sigma_{c_2}} = -\sqrt{\frac{Q_{c_1}}{1-Q_{c_1}} \frac{Q_{c_2}}{1-Q_{c_2}}} . \tag{B.10}$$

B.2 Computing the F's

For practical cases, it is not hard to compute the values of the $s_{c,i}$, and then the elements of $\mathbf{M}$. There exist canned routines that we can use to calculate the elements of the inverse of $\mathbf{M}$. We can then pick out those elements of $\mathbf{M}^{-1}$ that are the covariance matrix $\mathbf{V}$, and use the sigmas (the square roots of the diagonal elements) to calculate the significances $\alpha_c/\sigma(\alpha_c)$.

We know that the diagonal elements of $\mathbf{V}$ are $\frac{1}{N}Q_c(1-Q_c)$ for the case of complete separation, therefore we know that our calculated elements are $\frac{1}{NF}Q_c(1-Q_c)$, and we know what $\frac{1}{N}Q_c(1-Q_c)$ is, so we can easily obtain $\sqrt{1/F}$.

Bibliography

[1] C. Quigg. *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, pages 148–156. Benjamin/Cummings, 1988.

[2] Particle Data Group. *Phys. Rev.* **D45**, IX.12–17 (1992). Review of Particle Properties.

[3] N. A. Nicol. *Measurement of Tau Lepton Branching Fractions.* PhD thesis, University of California, Berkeley, September 1993. LBL-34784.

[4] P. F. Smith. *Ann. Rev. Nucl. and Part. Sci.* **39**, 73 (1989).

[5] C. A. Dominguez and E. de Rafael. *Ann. Phys.* **174**, 372 (1987).

[6] A. Zecca. *Phys. Lett.* **B213**, 210 (1988).

[7] F. Abe et al. *Evidence for Top Quark Production in $\bar{p}p$ Collisions at $\sqrt{s}$ = 1.8 TeV.* CDF Note 2561 FERMILAB Pub-94-097-E, Fermilab, Apr 1994. Submitted to *Physical Review D.*

[8] F. J. Yndurain. *Quantum Chromodynamics: an Introduction to the Theory of Quarks and Gluons.* Springer, 1983.

[9] Particle Data Group. *Phys. Rev.* **D45**, II.12–18,II.32–33 (1992).

[10] F. Abe et al. *Phys. Rev. Lett.* **71**, 1685 (1993).

[11] D. Buskulic et al. *Phys. Lett.* **B311**, 425 (1993).

[12] L. S. Brown. *Quantum Field Theory.* Cambridge University Press, 1992.

[13] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, page 228. John Wiley and Sons, 1984.

[14] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, page 171. John Wiley and Sons, 1984.

[15] T. Hebbeker. QCD Studies at LEP. In S. Hegarty, K. Potter, and E. Quercigh, editor, *Joint Int. Lepton Photon Symposium at High Energies and European Physical Society Conference on High Energy*, page 73, 1991.

[16] W. J. Stirling. Short Distance QCD. In W. Bartel and R. Rückl, editor, *Proceedings of the 1987 Int. Symposium on Lepton and Photon Interactions at High Energies, Hamburg*, page 715, 1988.

[17] S. G. Gorishny, A. Kataev, and S. A. Larin. *Phys. Lett.* **B259**, 114 (1991).

[18] L. R. Surguladze and M. A. Samuel. *Phys. Rev. Lett.* **66**, 560 (1991).

[19] C. Quigg. *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, pages 83–192. Benjamin/Cummings, 1988.

[20] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, pages 238–239. John Wiley and Sons, 1984.

[21] R. D. Field. *Applications of Perturbative QCD*, pages 57–106. Addison-Wesley, 1989.

[22] G. Martinelli. *Nucl. Phys.* **A527**, 89 (1991).

[23] R. Petronzio. Lattice Gauge Theories. In J. R. Sanford, editor, *Proc. of the 26th International Conference in High Energy Physics*, page 241, 1992.

[24] R. D. Field and R. P. Feynman. *Nucl. Phys.* **B136**, 1 (1978).

[25] G. D. Cowan. *Inclusive $\pi^{\pm}$, $K^{\pm}$, and p, $\overline{p}$ Production in e^+e^- Annihilation at $\sqrt{s}$ = 29 GeV.* PhD thesis, University of California, Berkeley, January 1988. LBL-24715.

[26] T. Meyer. *Z. Phys.* **C12**, 77 (1982).

[27] P. Hoyer et al. *Nucl. Phys.* **B161**, 349 (1979).

[28] A. Ali et al. *Phys. Lett.* **93B**, 155 (1980).

[29] B. Andersson et al. *Phys. Lett.* **94B**, 211 (1980).

[30] W. Bartel et al. *Z. Phys.* **C21**, 37 (1983).

[31] H. Aihara et al. *Phys. Rev. Lett.* **54**, 270 (1985).

[32] B. Foster. *Electron-Positron Annihilation Physics*, pages 62–64. Adam Hilger, 1990.

[33] X. Artru and G. Mennessier. *Nuc. Phys.* **B70**, 93 (1974).

[34] B. Andersson and G. Gustafson. *Z. Phys.* **C3**, 223 (1980).

[35] B. Andersson et al. *Nucl. Phys* **B135**, 273 (1978).

[36] B. Andersson et al. *Phys. Rep.* **97**, 31 (1983).

[37] D. H. Perkins. *Introduction to High Energy Physics*, pages 175–177. Addison-Wesley, 2nd (rev.) edition, 1982.

[38] J. Kogut and L. Susskind. *Phys. Rev.* **D9**, 3501 (1974).

[39] E. Eichten et al. *Phys. Rev.* **D17**, 3090 (1978).

[40] B. Andersson et al. *Z. Phys.* **C20**, 317 (1983).

[41] A. Ali and P. Söding. *High Energy Electron-Positron Physics*, pages 599–601. World Scientific, 1988.

[42] D. Amati and G. Veneziano. *Phys. Lett.* **83B**, 87 (1979).

[43] S. Wolfram. Parton and Hadron Production in e^+e^- Annihilation. In J. Tran Thanh Van, editor, *Proc. 15th Rencontre de Moriond*, page 549, 1980.

[44] B. R. Webber. *Nucl. Phys.* **B238**, 492 (1984).

[45] B. Foster. *Electron-Positron Annihilation Physics*, page 99. Adam Hilger, 1990.

[46] T. Sjöstrand, private communication.

[47] C. Peterson et al. *Phys. Rev.* **D27**, 105 (1983).

[48] D. Bortoletto et al. *Phys. Rev.* **D37**, 1719 (1988).

[49] J. Chrin. *Z. Phys.* **C36**, 163 (1987).

[50] M. Suzuki. *Phys. Lett.* **B71**, 139 (1977).

[51] J. D. Bjorken. *Phys. Rev.* **D17**, 171 (1978).

[52] B. Foster. *Electron-Positron Annihilation Physics*, pages 74–76. Adam Hilger, 1990.

[53] M. Althoff et al. *Phys. Lett.* **B135**, 243 (1984).

[54] R. Giles et al. *Phys. Rev.* **D30**, 2279 (1984).

[55] H. Albrecht et al. *Z. Phys.* **C54**, 13 (1992).

[56] P. Kesten et al. *Phys. Lett.* **B161**, 412 (1985).

[57] M. Sakuda et al. *Phys. Lett.* **B152**, 399 (1985).

[58] P. Rowson et al. *Phys. Rev. Lett.* **54**, 2580 (1985).

[59] H. Aihara et al. *Phys. Lett.* **B184**, 299 (1987).

[60] Particle Data Group. *Phys. Rev.* **D45**, II.17 (1992).

[61] R. Ong. *Measurement of the B Hadron Lifetime.* PhD thesis, Stanford University, September 1987. SLAC-320.

[62] Particle Data Group. *Phys. Rev.* **D45**, II.12–13 (1992).

[63] D. W. Ruck et al. *IEEE Transactions On Neural Networks* **1**, 296 (1990).

[64] L. S. Brown. *Quantum Field Theory*, page 442. Cambridge University Press, 1992.

[65] F. A. Berends and R. Kleiss. *Nuc. Phys.* **B178**, 141 (1981).

[66] F. A. Berends, R. Kleiss, and S. Jadach. *Computer Phys. Comm.* **29**, 185 (1983).

[67] J. J. Eastman. *Kaon Content of Three-Prong Decays of the Tau Lepton.* PhD thesis, University of California, Berkeley, December 1990. LBL-30035.

[68] H. Aihara et al. *Charged Hadron Production in e^+e^- Annihilation at $\sqrt{s}$ = 29 GeV.* LBL Report LBL-23737, Lawrence Berkeley Laboratory, March 1988.

[69] W. Gorn et al. *IEEE Trans. Nucl. Sci.* **NS-26**, 68 (1979).

[70] W. Gorn et al. *IEEE Trans. Nucl. Sci.* **NS-30**, 153 (1983).

[71] H. Aihara et al. *IEEE Trans. Nucl. Sci.* **NS-30**, 63, 76, 162 (1983).

[72] H. Aihara et al. *Nucl. Instr. Meth.* **223**, 40 (1984).

[73] H. Aihara et al. *IEEE Trans. Nucl. Sci.* **NS-30**, 117 (1983).

[74] H. Aihara et al. *Nucl. Instr. Meth.* **217**, 259 (1983).

[75] H. Aihara et al. *IEEE Trans. Nucl. Sci.* **NS-30**, 67 (1983).

[76] R. I. Koda. *A Test of Quantum Electrodynamics at Small Angles Using the PEP-4 Facility.* PhD thesis, University of California, Los Angeles, 1985. UCLA-85-011.

[77] M. P. Cain et al. *Phys. Lett.* **B147**, 232 (1984).

[78] J. W. Gary. *Tests of Models for Parton Fragmentation in e^+e^- Annihilation.* PhD thesis, University of California, Berkeley, November 1985. LBL-20638.

[79] M. T. Ronan. *The PEP-4 (TPC) Trigger System.* TPC Note TPC-LBL-87-12, Lawrence Berkeley Laboratory, May 1987.

[80] L. G. Mathis. *A Search for New Leptons with Heavy Neutrinos in e^+e^- Annihilation at $\sqrt{s} = 29$ GeV.* PhD thesis, University of California, Berkeley, May 1988. LBL-25261.

[81] T. K. Edberg. *Inclusive Production of Vector Mesons in e^+e^- Annihilation at $\sqrt{s} = 29$ GeV.* PhD thesis, University of California, Berkeley, August 1988. LBL-25652.

[82] T. K. Edberg. *A Neophyte's Guide to Track Fitting in the TPC.* TPC Note TPC-LBL-87-16, Lawrence Berkeley Laboratory, Feb 1987.

[83] J. D. Jackson. *Classical Electrodynamics*, pages 618–653. John Wiley and Sons, 1975.

[84] E. Fermi. *Phys. Rev.* **57**, 485 (1940).

[85] A. Crispin and G. Fowler. *Rev. Mod. Phys.* **42**, 290 (1970).

[86] D. Lambert. *dE/dx Corrections for Experiments 19-20, and dE/dx Resolution for Experiments 14-20.* TPC Note TPC-LBL-91-01, Lawrence Berkeley Laboratory, Jul 1991.

[87] H. S. Kaye. *PEP4 Offline Processing of Experiments 14–18.* TPC Note TPC-LBL-87-1, Lawrence Berkeley Laboratory, Jan 1987.

[88] Particle Data Group. *Phys. Rev.* **D45**, II.3 (1992).

[89] H. Yamamoto and N. Toge. *Estimates of Backgrounds in Hadronic Event Samples.* TPC Note TPC-UT-84-2, University of Tokyo, Jan 1984.

[90] J. Y. Oyang and C. D. Buchanan, private communication.

[91] B. Foster. *Electron-Positron Annihilation Physics*, page 67. Adam Hilger, 1990.

[92] J. Chrin, private communication.

[93] C. M. Bishop. *Rev. Sci. Instrum.* **65**, 1803 (1994).

[94] R. Lippman. *IEEE ASSP Magazine* **4**, 1 (1987).

[95] J. de Villiers and E. Barnard. *IEEE Transactions On Neural Networks* **4**, 136 (1992).

[96] L. Lönnblad, C. Peterson, and T. Rögnvaldsson. *Pattern Recognition in High Energy Physics with Artificial Neural Networks – Jetnet 2.0. Comp. Phys. Comm.* **70**, 167 (1992).

[97] K. H. Becks et al. *Nucl. Inst. Meth.* **A329**, 501 (1993).

[98] D.E. Rumelhart and J.L. McClelland (eds.). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 1).* MIT Press, 1986.

[99] T. D. Gottschalk and R. Nolty. *Identification Of Physics Processes Using Neural Network Classifiers.* Technical Report CALT-68-1680, California Institute of Technology, Nov 1990.

[100] R. Barlow. *J. Phys. G: Nucl. Part. Phys.* **17**, 1519 (1991).

[101] G. Bahan and R. Barlow. *Identification of b Jets using Neural Networks.* Technical Report MAN-HEP-92-01, Manchester University, Apr 1992.

[102] L. Lönnblad, C. Peterson, and T. Rögnvaldsson. *Nucl. Phys.* **B349**, 675 (1991).

[103] F. Seidel et al. B-quark tagging using neural networks and comparison with a classical method. In D. Perret-Gallix, editor, *2nd Int'l Workshop on Software Engineering, A.I., and Expert Systems for High-Energy and Nuc. Phys.*, page 425, 1992.

[104] L. Bellantoni et al. *Nucl. Instr. Meth.* **A310**, 618 (1991).

[105] C. Bortolotto, A. De Angelis, and L. Lanceri. *Nucl. Instr. Meth.* **A306**, 459 (1991).

[106] C. Peterson. *Neural Networks and High Energy Physics.* Technical Report LU-TP-90-6, University of Lund, Sweden, May 1990.

[107] G. Marchesini, G. Nardulli, and G. Pasquariello. *QCD Coherence in Tagging b Jets by Neural Networks.* Technical Report BARI-TH-92-98, I.N.F.N. Sezione di Parma, Italy, Mar 1992.

[108] B. Denby. Tutorial on Neural Networks Applications in High Energy Physics: A 1992 Perspective. In D. Perret-Gallix, editor, *2nd Int'l Workshop on Software Engineering, A.I., and Expert Systems for High-Energy and Nuc. Phys.*, 1992.

[109] B. Denby. *The Use of Neural Networks in High Energy Physics.* Technical Report Fermilab-PUB-92-215-E, Fermilab, Aug 1992.

[110] B. Denby. *Neural Networks and Cellular Automata in Experimental High Energy Physics. Comp. Phys. Comm.* **49**, 429 (1988).

[111] D. Cutts et al. The Use of Neural Networks in the D0 Data Acquisition System. In *Proc. Real Time '89*, 1989.

[112] Babage et al. *Nucl. Inst. and Meth.* **A330**, 482 (1993).

[113] V. Innocente, Y. F. Wang, and Z. P. Zhang. *Nucl. Inst. and Meth.* **A330**, 482 (1993).

[114] D. Buskulic et al. *Phys. Lett.* **B313**, 549 (1993). ALEPH Collab.

[115] O. Adriani et al. *Phys. Lett.* **B307**, 237 (1993). L3 Collab.

[116] P. Abreu et al. *Phys. Lett.* **B295**, 383 (1992). DELPHI Collab.

[117] P. Eberhard, G. Lynch, and D. Lambert. *Nucl. Inst. Meth.* **A326**, 574–580 (1993).

[118] F. James and M. Roos. *Comp. Phys. Comm.* **10**, 343 (1975).

[119] W. Hofmann. *Things You Always Wanted to Know, but Were Afraid To Ask.* TPC Note TPC-LBL-83-3, Lawrence Berkeley Laboratory, 1983.

[120] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics, Volume 1*, page 236. Hafner Press, 1976.

[121] W. T. Eadie et al. *Statistical Methods in Experimental Physics*, page 160. North Holland, 1971.

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720