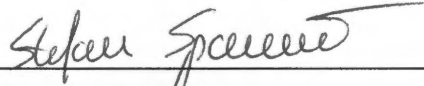


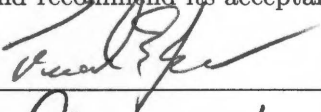
To the Graduate Council:

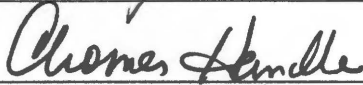
I am submitting herewith a thesis written by Gerald Conrad Ragghianti Jr. entitled "Analysis of a Gluonic Penguin Decay with the BaBar Detector." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Physics.



Stefan M. Spanier, Major Professor

We have read this thesis
and recommend its acceptance:





Accepted for the Council:



Interim Dean of Graduate Studies

ANALYSIS OF A GLUONIC PENGUIN DECAY WITH THE BABAR DETECTOR

A Thesis
Presented for the
Master of Science Degree
The University of Tennessee, Knoxville

Gerald Conrad Ragghianti Jr.
May 2007

Acknowledgments

First of all, I have to thank my parents Gerald and Sharon Ragghianti for providing an environment that allowed me to thrive and develop into the person I am today. Their support was instrumental in helping me find my own way in life.

In my work at The University of Tennessee, no one has been as instrumental in guiding my education as my adviser Dr. Stefan Spanier. He has showed me through the details of high energy physics, collaborative science, computing, and scientific writing. The many hours he has spent with me in front of a white board or in editing this thesis has been a testament to his dedication to science education. It has been an honor to work with a scientist of his caliber who is also willing to take time for his students. I also owe much to the tutelage of Dr. Mahalaxmi Krishnamurthy. Her research experience made for a very advantageous shoulder to stand on, and she was always willing to discuss, listen, and help. Furthermore, I would like to thank the other members of my committee: Dr. Thomas Handler and Dr. Ted Barnes. Their work and input into this thesis is greatly appreciated.

I also want to acknowledge the work of the nearly 600 researchers that make up the *BABAR* Collaboration. It has been a great experience to work within such a professional organization and in such leading scientific research. Without the work of *BABAR*, very little of this thesis would have been possible. In particular, I have greatly benefited from the help of the members of the Simulation Production group in my work there. Finally, Brad Gardner and Eugene McGuire of the physics department electronic shop were instrumental in the design and assembly of the computing cluster, and their support of our sometimes unconventional ideas and problems was very important for our success.

Abstract

This thesis presents a branching fraction analysis of the neutral B meson decay channel $B \rightarrow \phi K_S^0$ where the K_S^0 decays to $\pi^0\pi^0$. The decay is dominated by gluonic penguin transitions, which have been very important for the main program of *BABAR*: the search for physics beyond the Standard Model. The decay channel has been established and is included in the CP analysis, which is sensitive to new physics. The data set consists of 227 million $B\bar{B}$ pairs recorded by the *BABAR* detector at the Stanford Linear Accelerator Center. Sophisticated analysis techniques have been applied primarily to suppress background from $e^+e^- \rightarrow$ quark/anti-quark reactions. The analysis of such rare decay channels with *BABAR* relies on the availability of a large set of computer simulated data. For that purpose a computer cluster has been built at the University of Tennessee as part of the distributed computing support work for *BABAR*. The design and performance of the cluster is a main subject of this thesis work.

Contents

1	Physics Motivation	1
1.1	The Standard Model of Particle Physics	1
1.1.1	Fundamental Symmetries in the Standard Model	2
1.1.2	CP Violation and Matter-Anti-matter Asymmetry	3
1.2	CP Violation in the Standard Model	4
1.2.1	Quark Mixing	4
1.2.2	The Unitarity Triangle	6
1.3	Types of CP Violation	8
1.3.1	Penguin and Tree Amplitudes in B Decays	8
1.3.2	Mixing Amplitudes in B Decays	9
1.3.3	CP Violation in Decay	11
1.3.4	CP Violation in Mixing	12
1.3.5	CP Violation in Interference Between Mixing and Decay	12
1.4	Decay Rate Measurement	14
2	The <i>BABAR</i> Experiment	15
2.1	PEP-II B Factory	15
2.2	The <i>BABAR</i> Detector	18
2.2.1	Charged Particle Tracking	20
2.2.2	Particle Identification	23
2.2.3	Electromagnetic Calorimetry	27
2.2.4	Instrumented Flux Return	31
2.2.5	Summary	31

3	<i>BABAR</i> Computing Infrastructure	33
3.1	<i>BABAR</i> Monte-Carlo Simulation Production	33
3.1.1	The Monte-Carlo Method	33
3.1.2	The <i>BABAR</i> MC Simulation	34
3.1.3	Use of MC in <i>BABAR</i> Analysis	35
3.1.4	The Simulation Production Effort	35
3.2	Computing System Design	36
3.2.1	Diskless Compute Cluster	38
3.2.2	Compute Node Configuration	39
3.2.3	Central Servers	42
3.2.4	Network	44
3.2.5	Electrical Power	45
3.2.6	Air Conditioning System	45
3.3	Software Environment	46
3.3.1	Operating System	46
3.3.2	The <i>BABAR</i> Software Framework	47
3.3.3	Objectivity Database	48
3.3.4	ROOT Framework	48
3.3.5	Other Software Packages	49
3.3.6	File System Structure	49
3.3.7	<i>BABAR</i> Production Software	50
3.3.8	Custom Management Programs	50
3.4	Results	52
3.5	Future Expansion	54
4	Particle Reconstruction	57
4.1	Charged Particle Reconstruction	57
4.2	Particle Identification (PID)	58
4.3	Photon Reconstruction	59
4.3.1	π^0 Reconstruction	60
4.4	Event Reconstruction	60
4.4.1	ϕ Selection	60
4.4.2	K_S^0 Selection	63

5	Backgrounds	65
5.1	Continuum Background Suppression	65
5.2	B Meson Decay Backgrounds	65
6	Event Selection	69
6.1	Event Variables	69
6.1.1	m_{ES}	69
6.1.2	ΔE	69
6.1.3	Helicity	70
6.1.4	Fisher Discriminant	70
6.2	Final Event Selection	73
6.2.1	Multiple Candidates	73
6.2.2	Photon Association Study	73
6.3	Results	74
7	Use of Multi-layer Perceptron for Background Suppression	77
7.1	The Multi-layer Perceptron	77
7.2	Comparison with the Fisher Discriminant	80
7.3	Network Topology	81
7.4	Network Training	81
7.4.1	Training Methods	82
7.4.2	Training Procedure	82
7.5	Results	83
7.5.1	$B^0 \rightarrow \phi K_S^0, K_S^0 \rightarrow \pi^0 \pi^0$ Dataset	83
7.5.2	$B^0 \rightarrow \phi K_L^0$ Dataset	83
7.6	Conclusion	87
8	Analysis	89
8.1	Cut-and-Count Analysis	89
8.1.1	Optimized Cuts	90
8.1.2	Counting Analysis	90
8.1.3	B Background	97
8.1.4	Box-counting Cross-check	97
8.2	Maximum Likelihood Fit to Event Yields	100
8.2.1	Event Variable Parameterization	101
8.2.2	Fit Stability Analysis	102
8.2.3	Systematic Uncertainties	107

8.2.4	Result	109
8.3	<i>CP</i> Analysis	113
8.3.1	Introduction	113
8.3.2	Strategy	113
8.3.3	Measurement of Δt	115
8.3.4	<i>B</i> Flavor Tagging	117
8.3.5	<i>B</i> Flavor Sample	118
8.3.6	Maximum Likelihood Fit	119
8.3.7	Studies of the Likelihood Fit	119
8.3.8	Systematic Error	120
8.3.9	Result	120
8.4	Conclusion	121
	Bibliography	123
	Vita	133

List of Tables

2.1	Production cross-section at the $\Upsilon(4S)$ for fermion pairs	16
2.2	Interaction processes of photons with material in different regimes of the photon energy E	28
2.3	Overview of the angular coverage (in the <i>BABAR</i> center-of-mass frame) for each detector sub-system	31
5.1	The B background estimate from Monte-Carlo data sets shows a very small background contribution from generic B decays. Estimates of the signal yield in data are made based on either a sample of signal MC (first row) or the sample of $B^0\bar{B}^0$ generic MC (row four). The difference in these two values is attributed to the much higher uncertainty in the second value due to smaller sample size.	66
6.1	Correlation coefficients for each combination of event shape variables used for continuum background suppression	72
6.2	We used 578,000 signal MC events to determine the fraction of mis-association of π^0 photons. We find that 15.3% of events have photons that are incorrectly associated with the π^0 parent, are missing one or more photons, or have other reconstruction errors that preclude correctly assigning the π^0 photons.	75
6.3	This analysis of pre-selection and event selection requirement efficiencies uses signal Monte-Carlo. All efficiencies following pre-selection are with respect to the number of events passing pre-selection.	76

7.1	This is an investigation of the dependence of the final MLP output Error on network size with the objective of choosing a network which results in the lowest Error after a set number of training epochs (10,000). All networks used a single hidden layer. σ_{Error} was calculated from ten separate training attempts for each trial. For this data sample ($B^0 \rightarrow \phi K_s^0$, $K_s^0 \rightarrow \pi^0 \pi^0$), there appears to be little advantage in using a network having more hidden nodes than number of input variables.	86
7.2	This is an investigation of the effect of training algorithm and MLP network size on training speed and final MLP Error. Each network was sufficiently trained to ensure convergence of the Error to a minimum. Training algorithms are detailed in Ref. [1]	88
8.1	The three regions used for the cut-and-count analysis: ΔE sideband, Grand Sideband, and the signal region	92
8.2	These correlation coefficients between the event variables (from signal Monte-Carlo and sideband on-resonance background events) include the measured proper decay time difference Δt between the two B mesons in the event and the error $\sigma_{\Delta t}$. The variables will be explained in detail in Section 8.3.2.	94
8.3	Results from the Cut and Count cross-check analysis for signal yield . . .	98
8.4	Parameterization of the event variable PDFs for the final state $B^0 \rightarrow \phi K_s(\pi^0 \pi^0)$ (kinematic quantities in GeV)	103
8.5	The systematic error due to signal PDF parameters is evaluated by calculating the relative shift of the number of signal events N_{sig} due to a shift in each PDF parameter by ± 1 standard deviation. The component errors were added in quadrature for the total PDF parameter systematic error.	108
8.6	Estimate of the total yield fit systematic error sources	109
8.7	Results from the branching fraction analysis	110

List of Figures

1.1	Weak charged-current couplings with 3 quark generations	6
1.2	The Unitarity Triangle where all sides have been divided by $V_{cb}^* V_{cd}$	7
1.3	Tree diagram (left) and penguin diagram (right) for the neutral, $B^0(\bar{b}d)$, and charged, $B^+(\bar{b}u)$, meson decay	8
1.4	Penguin diagram involving non-SM quarks and mediator particles as predicted by supersymmetric theories	10
1.5	Two mixing (or “box”) diagrams where the quarks $\bar{q}_1$ and q_2 can have either d , s or b flavor	10
2.1	At the SLAC PEP-II accelerator facility, electrons and positrons are accelerated to high energies in the linear accelerator (not shown) and injected into two opposing storage rings.	17
2.2	B Factory integrated luminosity since 1999 (start of physics data taking) until the end of this analysis showing the luminosity provided by PEP-II (darkest line), recorded by <i>BABAR</i> ($\approx 96\%$ efficiency) (middle line), and recorded at 40MeV below the $\Upsilon(4S)$ resonance (lightest line)	19
2.3	Cutaway diagram of the <i>BABAR</i> detector that shows the Silicon Vertex Tracker (1), Drift Chamber (2), Cherenkov Detector (3), Electromagnetic Calorimeter (4), superconducting solenoid providing 1.5 Tesla magnetic field (5), and Instrumented Flux Return (6)	21
2.4	Details of the inner parts of the detector, showing the cross-sectional view of the silicon vertex tracker (SVT) in a plane containing the beam axis	21
2.5	Side view of the <i>BABAR</i> drift chamber (DCH) with units in mm	23
2.6	This schematic view of one DIRC quartz bar shows how the Cherenkov light cone is reflected within the quartz bar and guided to the readout PMTs within the water-filled standoff box.	24

2.7	The measured Cherenkov angle of tracks from an inclusive sample of multi-hadron events plotted against the momentum of the tracks with gray lines indicating the predicted values of θ_C for different particle species.	26
2.8	dE/dx information from the drift chamber plotted as a function of particle momentum shows clear differentiation between particle species.	28
2.9	A side view of one line of EMC sensors showing both the barrel section and forward end-cap (to account for the boosted center-of-mass frame) with units in mm	30
3.1	This is a schematic of the compute cluster configuration with the two central file servers connected to the network switches via 1 Gbit/sec uplink ports. The switches are connected together and then connected to the compute nodes via 100 Mbit/sec ports. The compute nodes are separated into two groups, each connected to one of the network switches.	37
3.2	The compute nodes only contain components essential for operation: the motherboard, CPUs, RAM, network card, two fans, and a PC power supply.	40
3.3	The full rack of 40 compute nodes that make up the 80 CPU cluster consist of machines based on Athlon MP 2000+ (left) and Athlon MP 2400+ (right) CPUs. Two network switches are located above the second level of nodes. The nodes are mounted facing the rear of the rack in order to facilitate access to the network and power connections and to exhaust waste heat into the center of the room.	43
3.4	The online cluster monitoring software features a graph display of each compute node that shows the CPU load, internal temperatures, and CPU fan speed. Clicking on an individual node will show a report of CPU use, temperature, network activity, and virtual memory use over the last 24 hours. The status fields on the overall cluster view are color coded in order to quickly provide an indication of the activity of the cluster and to highlight problem areas.	51

3.5	The Tennessee compute cluster has produced 250 million MC simulated events as of January 2005. This plots shows an upward kink in the MC production rate at about week 30 caused by the expansion of the cluster (which roughly doubled the available computing power). This increased production capacity was tempered somewhat by a software version upgrade (SP5 to SP6) that introduced a more computationally intensive simulation.	53
3.6	This shows the total proportion of MC events produced per <i>BABAR</i> SP site as of Jan. 2005. The contribution from the University of Tennessee <i>BABAR</i> group represents nearly 10% of total events produced and ranks our site among the top 7 university-based production sites.	55
4.1	This shows the identification efficiency (left) of the charged kaon selection as a function of the kaon laboratory momentum p_K^{lab} and the fraction of charged pions misidentified (mis-ID) as kaons as a function of the pion laboratory momentum p_π^{lab} (right). The error bars represent statistical uncertainties in the control sample of kaons and pions from $D^* \rightarrow D^0\pi$, $D^0 \rightarrow K\pi$ decays. It is not required that the particle track traverses all particle identification detectors; notice that the DIRC solid angle coverage is only 87%.	62
5.1	This diagram shows the general distribution of decay tracks from $B\bar{B}$ events (a) and continuum background events (b) in the (r,ϕ) CM frame. Continuum background event particles travel in <i>jets</i> which group around opposing axes (the thrust axes) while $B\bar{B}$ events are isotropically distributed in ϕ	66
6.1	Distributions of the four topological variables used in continuum background suppression for signal (solid red) and background (dashed blue) including the Thrust angle (a), B production angle (b), $\mathcal{L}_0$ (c), and $\mathcal{L}_2$ (d) with arbitrarily normalization.	71
6.2	Signal Monte-Carlo shows significant differences in m_{ES} (left) and ΔE (right) between well-reconstructed signal events (solid) and SXF signal (error bars). Both histograms for SXF and good signal are normalized to the same number of events.	74

7.1	This diagram of a Multi-layer Perceptron (MLP) consists of an input, an output, and one hidden layer. Each layer contains one or more nodes. Each node is connected to every other node in adjacent layers by lines to represent the weights used in the MLP calculation. The example MLP network consists of three input nodes, two hidden layer nodes, and one output node. Equation 7.4 shows the equivalent mathematical notation for this specific MLP network.	78
7.2	This plot shows the progress of the MLP network Error (Equation 7.2) as a function of training epoch for three different network sizes. There is an initial downward trend for all errors. However, larger networks can become over-trained and eventually result in an increased error after a number of epochs (as seen in the 3-20-1 network after epoch 700). The training should be halted when the error reaches a minimum.	84
7.3	This shows the distribution of the MLP output for background (shaded) and signal (clear) events in an optimized background separation MLP consisting of four input nodes, four hidden layer nodes, and one output node.	85
8.1	The plot on the left shows the signal significance versus range cutoff for the Fisher event variable. The right plot shows the location of the optimized cut (vertical line) (maximum signal significance) on the variable's distributions for signal MC (solid) and sideband on-resonance background (dashed) samples.	91
8.2	The plot on the left shows the signal significance versus cutoff for the Helicity event variable ($ \cos\theta_H $). The right plot shows the location of the optimized cut (vertical line) (maximum signal significance) on the variable's distributions for signal MC (solid) and sideband on-resonance background (dashed) samples.	91
8.3	These plots show the distributions of signal MC events in the signal box interval for m_{ES} (left) and ΔE (right).	92
8.4	The regions in ΔE and m_{ES} listed in Table 8.1 are identified on the plot: ΔE sideband (a, b, e, and f), Grand Sideband (a, c and e), and the signal region (d).	93
8.5	The ΔE distribution in data from the GSB (dots) with the fit to a linear function superimposed and dotted lines indicating the 1σ range of uncertainty in the fitted slope value	95

8.6	The m_{ES} distribution is parameterized with an “Argus” shape function (Equation 8.2) and is fit to data events in the ΔE sidebands (ranges $\Delta E < -0.012 \text{ GeV}$ and $\Delta E > 0.056 \text{ GeV}$) (left). The dotted lines show the $\pm 1\sigma$ ranges of uncertainty in the Argus shape parameter. We fit m_{ES} to data in the Grand Sideband (GSB) region ($m_{\text{ES}} < 5.265 \text{ GeV}/c$) (right) as a cross check on the shape parameter. Both parameter sets are in good agreement.	96
8.7	We calculate the number of background events expected in the m_{ES} signal region (right shaded area) by scaling the number of events counted in the GSB (left shaded area) by the ratio of areas under the fitted Argus shape function.	96
8.8	This plot of post-cut experimental data shows the signal region (cross-hatched). The region $5.26 \text{ GeV}/c^2 < m_{\text{ES}} < 5.27 \text{ GeV}/c^2$ (diagonal lines) between the sideband region and the signal region is ignored for background estimates.	99
8.9	The Four event variable parameterizations for well-reconstructed signal events fitted on signal MC data with an arbitrary normalization	104
8.10	The parameterizations of the four event variables for SXF signal events fitted on SXF signal MC with arbitrary normalization	105
8.11	The left figure shows the distribution of our result N_{sig} from fits to 2000 toy data set. The dotted line represents our result. The right plot shows the “pull” distribution of the signal yield from these fits. The variable “pull” is explained in the text. Both histograms have an arbitrary normalization.	106
8.12	Histograms of per-event $\log(\text{likelihood})$ for toy MC (solid) and experiment data (points with error bars) show that the likelihood treatment is similar for both sets of events. Event yields have been normalized and are shown on a logarithmic scale.	106
8.13	The $-\ln L/L_{\text{max}}$ of the yield fit PDF on data as a function of the number of signal events shows the statistical uncertainty of the fit at $N_{\text{sig}} = 0.0$ events. This corresponds to a 5.6 standard deviation statistical significance for the signal yield measurement.	111
8.14	Likelihood-enhanced plots of the four event variables showing signal above the background level (dashed)	112
8.15	Schematic of the measurement of the time-dependent CP asymmetry in neutral B decays with the BABAR experiment	114

8.16 Plots (a) and (b) show the Δt distributions of B^0 and $\bar{B}^0$ -tagged ϕK_S^0 candidates. The solid lines refer to the fit for all events; the dashed lines correspond to the background. Plot (c) shows the raw asymmetry. Plots (d), (e), and (f) are the corresponding plots for ϕK_L^0 candidates. For each final state, a requirement is applied on the event likelihood to suppress background.	122
--	-----

Chapter 1

Physics Motivation

1.1 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) is currently the most accepted model of fundamental particles and forces. It describes elementary particle reactions up to energies achieved by present-day accelerators to a very high precision [2]. It has been developed between 1970 and 1973 [3] and the formulation of its mathematical foundation was rewarded with a Noble prize in 1999 [4]. The basic ingredients for this quantum field theory are 12 flavors of fermions, 3 quark generations with fractional electrical charges (+2/3, -1/3), and 3 lepton generations with integer charges (-1, 0):

$$\text{quarks: } \begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \quad \text{leptons: } \begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}. \quad (1.1)$$

The forces between the fermions are modeled by coupling them to bosons, which mediate the forces. The Standard Model is primarily the theory of electroweak interactions, which describes the reactions invoking weak and electromagnetic forces, and combines it with the theory of hadronic interactions (quantum chromodynamics). The bosons in the SM are:

- Photons, which mediate the electromagnetic interaction,
- W^+ , W^- , and Z bosons, which mediate the weak nuclear force,
- Eight species of gluons, which mediate the strong nuclear force,
- The Higgs bosons, which induce spontaneous breaking of the electroweak symmetry and are responsible for the existence of inertial mass.

The Standard Model is not a complete theory of fundamental physics because it has two important defects:

1. The model contains 19 free parameters, such as particle masses, which must be determined experimentally (plus another 10 for neutrino masses). These parameters cannot be independently calculated.
2. The model does not describe gravitational interaction.

Since the completion of the Standard Model, many theories have been developed to address these problems [5, 6].

It has been recognized that the SM contains the ingredients to model the evolution of the Universe [7]. This work addresses one important aspect: In the SM, matter and anti-matter are related by the so-called *CPT* symmetry. This suggests that there should be equal amounts of matter and anti-matter after the Big Bang. While the dominance of matter in today's Universe can be explained by assuming that the Universe just started out this way, this strikes most physicists as inelegant.

1.1.1 Fundamental Symmetries in the Standard Model

In general, a symmetry of a physical system is characterized by a transformation that does not change the system or its appearance. The identification of fundamental symmetries in the Lagrangian of a theory allows a higher level of abstraction in its formulation. There is a one-to-one relation between an inherent symmetry of a system and a related conservation law [8]. There are fundamental requirements such as Lorentz invariance to a quantum field theory. We address some fundamental discrete symmetries of immediate relevance for this work below.

Parity

The parity P transformation is the reversal of the signs of all three spacial coordinates: $P(x, y, z) = (-x, -y, -z)$. This operation also reverses the sign of the particle's momentum while leaving its angular momentum (and spin) unchanged. Electromagnetic and strong interactions are invariant under parity transformation, but the weak interaction is not. Parity violation was first observed in β decays of Cobalt 60 by C.S. Wu [9] by measuring a bias in emission direction of electrons from a uniformly oriented Cobalt 60 sample.

Charge Conjugation

Charge conjugation C is the transformation of a particle Q into its anti-particle $\bar{Q}$: $CQ = \bar{Q}$. This transformation leaves the particle's mass, energy, momentum, and spin unchanged. Electromagnetism, gravity and the strong interaction obey C -symmetry, but weak interactions violate C -symmetry maximally. A simple example is charge-conjugation transformation of the left handed neutrino into the left-handed anti-neutrino, which has never been observed [10]. This indicates that any process involving neutrinos is not charge-conjugation symmetric.

Time Reversal

The time reversal T reverses the time-component: $T t = -t$. As a consequence, it reverses the momentum and spin of a particle.

Combined Transformations

The electromagnetic and strong force is invariant under C , P , and T , separately. However, the weak force violates C and P individually. Weak interactions were thought to conserve the combined symmetry CP until 1964 when CP violation was discovered in neutral kaon decays [11, 12]. The Standard Model allows for CP violation, as will be described below.

The CPT symmetry is a fundamental symmetry of physical laws under transformations that involve the inversions of charge, parity and time simultaneously. Any Lorentz invariant local quantum field theory with a Hermitian Lagrangian must have CPT symmetry. This is the *CPT theorem*, which was explicitly proven in 1954 by Gerhard Lüders [13], Wolfgang Pauli [14], and John Stewart Bell [15] and later by Res Jost [16] in a more general way. Because of the CPT -symmetry, a violation of the CP -symmetry is equivalent to a violation of the T -symmetry.

1.1.2 CP Violation and Matter-Anti-matter Asymmetry

The decay of a particle Q into the final state f is transformed by the CP operation into the decay of the corresponding anti-particle $\bar{Q}$ into the final state $\bar{f}$, and with the corresponding changes in the kinematics:

$$CP(Q \rightarrow f) = \bar{Q} \rightarrow \bar{f}. \quad (1.2)$$

CP symmetry implies that the decay rates for particle and anti-particle are identical. On the other hand, CP violation will cause a difference in the rates and hence is a

natural ingredient in the explanation of the matter anti-matter asymmetry in the Universe, assuming it started out symmetric. Shortly after the discovery of CP violation A. Sakharov [17] formulated three conditions that allow a link between the microscopic physics of the SM with cosmological models describing baryon¹ generation:

1. Violation of the quark number, i.e., the sum of all quarks (anti-quarks) before a reaction is not the same after the reaction; this condition is also known as baryon-number violation.
2. C and CP symmetry violation causing different production rates for quarks over anti-quarks (baryons over anti-baryons)
3. The baryon-number violating processes must be out of equilibrium. At some point in time the baryon asymmetric phase was frozen in time.

There is no current experimental evidence of particle interactions that violate the quark number. The CP violation in the SM has been measured by *BABAR* in the neutral B meson system with very high precision [18]. However, the allowed and measured CP violation in the SM appears far too small for all cosmological models constructed so far [19]. Those facts suggest the existence of physics beyond the SM.

1.2 CP Violation in the Standard Model

The parameterization of CP violation and the number of quark flavors are tightly coupled [20], as will be explained below.

1.2.1 Quark Mixing

In weak interaction, quarks can transition from one flavor to another due to the exchange or emission of $W^\pm$ bosons. These quark states are mixed states of the mass eigenstates of the flavor-preserving strong interaction. Historically, when only two generations of quarks were known, the quark mass eigenstates (d and s) translated into electroweak eigenstates (d' and s') according to a 2×2 rotation or “mixing” matrix:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \quad (1.3)$$

All weak decays involving u, d, s and c quarks were found to be consistent with a unique value of the mixing angle θ_C known as the Cabibbo angle [21].

¹Baryons are particles composed of 3 quarks, including protons and neutrons.

In general, the rate for a transition from an initial state i to a final state f is proportional to the transition amplitude $\mathcal{M}_{fi}$ squared. For weak interactions this transition (decay) amplitude has the form:

$$\mathcal{M}_{fi} = \frac{g}{\sqrt{2}} J^\mu \frac{1}{M_W^2 - q^2} \frac{g}{\sqrt{2}} J_\mu^\dagger, \quad (1.4)$$

with J^μ the weak charged current, M_W and q the mass and momentum carried by the vector boson W^+ or W^- mediating the interaction, and g the dimensionless weak coupling constant. The factor $1/\sqrt{2}$ is inserted for the conventional form of g . The weak charged currents expressed in matrix form are:

$$J = (\bar{u}, \bar{c}) \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}, \quad (1.5)$$

omitting the space-time structure operator.

This mixing matrix was extended by Kobayashi and Maskawa [20] to three generations of quarks in what is commonly known as the Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix $\mathbf{V}_{\text{CKM}}$. The weak charged currents become:

$$J = (\bar{u}, \bar{c}, \bar{t}) \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (1.6)$$

The currents can be represented by the diagram shown in Figure 1.1. The CKM matrix is unitary ($VV^\dagger = 1$; $V^\dagger$ is the conjugate complex transposed matrix), accounting for the quark number conservation, and a standard parameterization [22] with three Euler angles and one non-trivial phase δ is:

$$\begin{pmatrix} c_{12}c_{13} & c_{13}s_{12} & s_{13}e^{-i\delta} \\ -c_{23}s_{12} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \quad (1.7)$$

with $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$. It is known experimentally [23] that $s_{13} \ll s_{23} \ll s_{12} \ll 1$. The phase δ introduces the possibility of a T or CP violating amplitude, since under time reversal T $e^{i\delta} = e^{-i\delta}$. In general, for n generations of quarks ($2n$ quark flavors) because of the unitarity of V , phase conventions and one unobservable common phase, the number of real parameters in V is $(n-1)^2$, of which $n(n-1)/2$ are mixing angles and the remaining $(n-1)(n-2)/2$ are complex phases which cause CP violation [24].

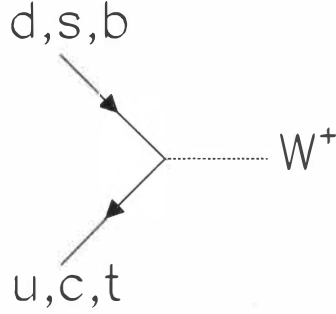


Figure 1.1: Weak charged-current couplings with 3 quark generations

1.2.2 The Unitarity Triangle

The unitarity of the CKM matrix imposes six relations among its columns or rows. The equation relevant for the study of CP violation in the B_d meson system is

$$V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0, \quad (1.8)$$

where the star indicates the complex conjugate element. The expression requires the sum of the three complex numbers to vanish and can be represented as a triangle in the complex plane. The area of the triangles for all six sums is the same and is directly proportional to the amount of CP violation [24]. By dividing each side by the best known $V_{cb}^* V_{cd}$ the vertices are exactly at $(0, 0)$, $(1, 0)$, and $(\bar{\rho}, \bar{\eta})$ ($\bar{\rho} + i\bar{\eta} = -(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)$). The triangle is displayed in Figure 1.2.

The three angles of the unitarity triangle are denoted by α, β, γ (or ϕ_1, ϕ_2, ϕ_3) [25] and are given by:

$$\begin{aligned} \alpha &\equiv \arg \left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right], \quad \beta \equiv \arg \left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right], \quad \text{and} \\ \gamma &\equiv \arg \left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right] = \pi - \alpha - \beta. \end{aligned} \quad (1.9)$$

These are physical quantities and can be measured by CP asymmetries in various B meson decays, e.g., α in $B^0 \rightarrow \pi^+ \pi^-$, β in $B^0 \rightarrow J/\psi K^0$, and γ in $B^0 \rightarrow DK$. The sides of the triangle are measurable from semi-leptonic branching fractions. By performing redundant measurements on the sides and angles one probes for violation of the consequences of unitarity that would imply the presence of other particles not included in the SM theory.

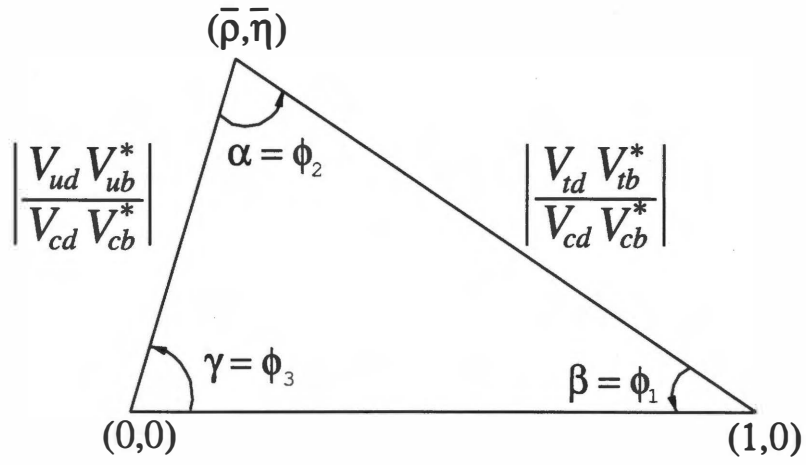


Figure 1.2: The Unitarity Triangle where all sides have been divided by $V_{cb}^* V_{cd}$

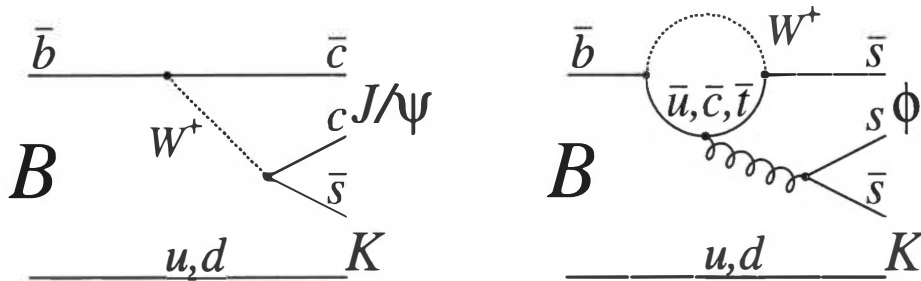


Figure 1.3: Tree diagram (left) and penguin diagram (right) for the neutral, $B^0(\bar{b}d)$, and charged, $B^+(\bar{b}u)$, meson decay

1.3 Types of CP Violation

The possible manifestations of CP violation can be described in a model independent way:

- CP violating in decays (direct CP violation),
- CP violation in mixing (indirect CP violation),
- CP violation in the interference between mixing and decay.

All types can be accessed with B_d meson decays. In general, at least two contributing transition amplitudes are required to project out the relative (weak) phase in rate measurements. The transition rate is proportional to the total amplitude squared,

$$R \propto |A|^2 \quad \text{with} \quad A = \sum_j a_j e^{i\phi_j}, \quad (1.10)$$

which is the sum over contributing amplitudes, with a_j being their magnitudes, ϕ_j their relative phases, and one unobservable overall phase. For one contribution only, $j = 1$, the rate is just $R \propto |a_1|^2$.

1.3.1 Penguin and Tree Amplitudes in B Decays

In the Standard Model, the relevant amplitudes for B_d decays are described by either the direct emission of a W boson (so-called tree decay) or a quark- W -boson loop (so-called penguin decay) as displayed in Figure 1.3. The other quark in the B_d meson (u or d) is assumed to be a spectator of the decay.

If a gluon is emitted from the loop in the penguin diagram, it is called *gluonic penguin*. While in an electroweak penguin, either a virtual photon or a Z^0 boson is emitted. In radiative penguins ($b \rightarrow s\gamma$) the emitted photon is real and detectable. In the gluonic penguin, the gluon turns into a quark-anti-quark pair. In Figure 1.3 (right), the quarks in the final state arrange themselves to constitute a $\phi(1020)$ meson (mostly $\bar{s}s$) and a charged or neutral kaon ($u\bar{s}$ or $d\bar{s}$). This is the final state studied in this thesis.

Loop decays in general are strongly suppressed, but they play a significant role for b -quarks because the tree decays are also suppressed, which explains the relatively long lifetime of the B mesons of about 1.5 ps. Hence, tree and penguin amplitudes are of comparable strength in many decays, e.g., $B^0 \rightarrow \pi^+\pi^-$. Since the penguin amplitudes are typically governed by different couplings, they often spoil the simple relationship between the fundamental phase of the SM and measurable CP asymmetries (penguin pollution). Therefore, a better understanding of the strength and nature of penguin decays is needed for clean interpretations of many CP measurements at *BABAR* and future experiments. The decay $B^0 \rightarrow J/\psi K^0$, though, is dominated by tree decays, and the next order penguin decay couplings have the same weak phase. In a variety of decays, such as $B^0 \rightarrow \phi K^0$, the tree contributions are strongly suppressed (proceeding almost exclusively through the penguin amplitude). In the SM, the CP asymmetries in the two channels $J/\psi K^0$ and ϕK^0 are sensitive to the same weak angle β .

Penguin dominated decays are of particular interest because they can signal the presence of new physics in an undisturbed way. New constituent particles and mediator particles beyond the SM can add to the virtual loop. If these new particle generations involve new mixing phases, they can change the angles of the unitarity triangle and cause differences in CP measurements with respect to SM expectations. Figure 1.4 shows a hypothetical penguin diagram that assumes the existence of supersymmetric particles.

Penguin amplitudes were first considered for B meson decays by J. Ellis et al. [26]. The name “penguin” originates from a game of darts in a pub in Geneva, Switzerland involving theorists John Ellis and others with the challenge that the loser of the game had to use the word penguin in his or her next paper [27].

1.3.2 Mixing Amplitudes in B Decays

From the *BABAR* Physics Book [28]: the four pairs of conjugate neutral mesons that decay weakly, K^0 , D^0 , B_d^0 , and B_s^0 , can each mix with their respective anti-particle via a pair of box diagrams shown in Figure 1.5. The ability to mix implies that the flavor eigenstates are not equivalent to the mass eigenstates. The light (L) and heavy (H) B

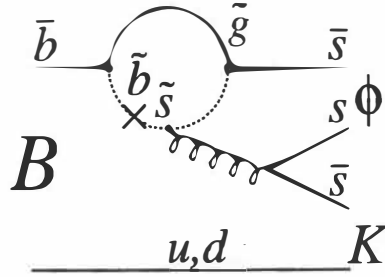


Figure 1.4: Penguin diagram involving non-SM quarks and mediator particles as predicted by supersymmetric theories

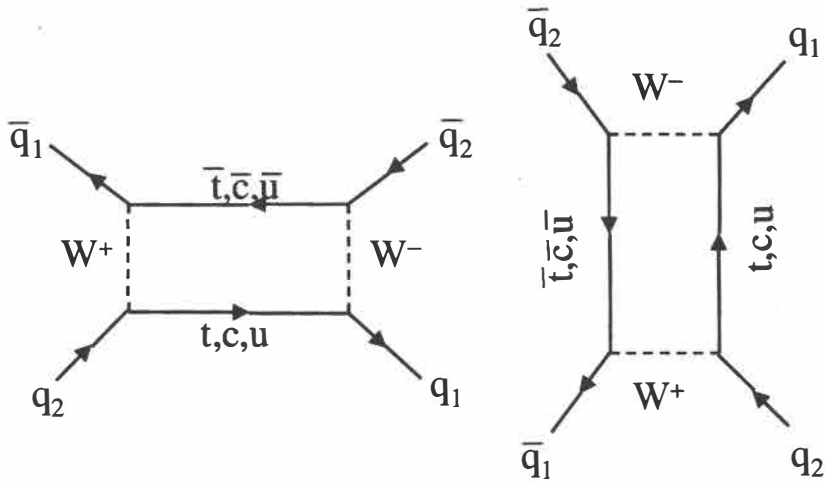


Figure 1.5: Two mixing (or "box") diagrams where the quarks $\bar{q}_1$ and q_2 can have either d , s or b flavor

mass eigenstates

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle, \quad \text{where } |p|^2 + |q|^2 = 1, \quad (1.11)$$

have a mass difference $\Delta m = m_H - m_L > 0$, and a total decay width difference $\Delta\Gamma = \Gamma_L - \Gamma + H$. As a consequence, an initially pure $|B^0\rangle$ state will evolve in time as a superposition of the mass eigenstates. The evolution of a pure $|B^0\rangle$ or $|\bar{B}^0\rangle$ state at the time $t = 0$ is given by:

$$\begin{aligned} |B^0(t)\rangle &= g_+(t)|B^0\rangle + \frac{q}{p}g_-(t)|\bar{B}^0\rangle \\ |\bar{B}^0(t)\rangle &= g_+(t)|\bar{B}^0\rangle + \frac{p}{q}g_-(t)|B^0\rangle, \end{aligned} \quad (1.12)$$

where g_+ and g_- are the amplitudes for the unchanged and changed flavor states, respectively, and $\Gamma = (\Gamma_H + \Gamma_L)/2$. The Δm has been measured by BABAR [29] to be 13 orders of magnitude smaller than the mass of the neutral B meson of $m = 5.29 \text{ GeV}/c^2$. Furthermore, for the B meson, $\Delta m \gg \Delta\Gamma$, and the corresponding time-dependent probabilities are

$$|g_{\pm}(t)|^2 = \pm \frac{e^{-\Gamma t}}{2} \cos(\Delta m t). \quad (1.13)$$

Mixing in the neutral B meson system is an essential ingredient for the measurement of theoretically clean time-dependent CP asymmetries (as explained below).

1.3.3 CP Violation in Decay

Direct CP violation is observed as a difference in the decay rate between a particle and its CP conjugate. The general form of the decay amplitude for B mesons in Eq. 1.10 consists of three components: the magnitude a_j , a weak phase ϕ_j , and a strong phase δ_j . The CP violating weak phase may arise from the SM CKM matrix or from new physics, while the CP conserving strong phase may arise from interactions between the hadronic particles in the final state:

$$A = \sum_j a_j e^{i\phi_j} e^{i\delta_j}; \quad \text{CP conjugate: } \bar{A} = \eta \sum_j a_j e^{-i\phi_j} e^{i\delta_j}, \quad (1.14)$$

where η is relevant if the final state is a CP eigenstate, and its value is ± 1 depending on the CP eigenvalue. The difference in yields becomes:

$$|A|^2 - |\bar{A}|^2 = -2 \sum_{j,k} a_j a_k \sin(\phi_j - \phi_k) \sin(\delta_j - \delta_k). \quad (1.15)$$

This means that direct CP violation will not cause a difference unless at least two contributing amplitudes have different weak phases *and* different strong phases. Precise predictions of rate differences or asymmetries a_{CP} , with

$$a_{CP} = \frac{|\bar{A}|^2 - |A|^2}{|\bar{A}|^2 + |A|^2}, \quad (1.16)$$

are not feasible at present as both the absolute value and the strong interaction phases of contributing amplitudes are not calculable from Lagrangians. However, numerical estimates can be done under well-defined model assumptions, and dependence on model and CKM parameters can be probed experimentally. Direct CP violation is the only kind that can also occur in the charged B system.

A sizable direct CP asymmetry has been measured by the *BABAR* collaboration in the decay $B^0 \rightarrow K^+\pi^-$, where the charge of the kaon tags the flavor of the neutral B meson. From a data sample of 227 million $\Upsilon(4S) \rightarrow \bar{B}B$ decays [30], (1606 ± 51) decays of this kind were reconstructed and the rate asymmetry corresponding to Eq. 1.16 is $a_{CP} = -0.133 \pm 0.030(\text{stat}) \pm 0.009(\text{syst})$, indicating a faster decay of the anti-particle $\bar{B}^0$.

1.3.4 CP Violation in Mixing

If in the description of mixing, Eq. 1.11, q and p have different magnitudes, then the CP conjugates of the mass eigenstates are different from the mass eigenstates by more than a trivial phase. Thus, the mass eigenstates are not CP eigenstates, and CP violation is present. It is purely an effect of mixing and is independent of decay mode.

This type of CP violation has been observed in the neutral kaon system [11]. The discovery of CP violation in 1964 was a detection of CP violation purely in mixing.

Effects of CP violation in mixing in the neutral B_d decays, such as asymmetries in B -flavor tagging semileptonic decays, are expected to be small ($\approx 10^{-3}$). At the present level of experimental precision, this CP violation is a negligible effect. Even if such asymmetries are observed, the dependence on large hadronic uncertainties makes it difficult to relate them to fundamental CKM parameters [23].

1.3.5 CP Violation in Interference Between Mixing and Decay

In the decays of $B^0 \rightarrow f$ and $\bar{B}^0 \rightarrow \bar{f}$, the final states, f and $\bar{f}$, that may be reached from either B^0 or $\bar{B}^0$ decays ($f = \bar{f}$) can exhibit a third type of CP violation resulting from the interference between the decays of mixed and unmixed neutral B mesons that both decay to the same final state.

For the neutral B system, such CP violation can be observed by comparing decay rates into final CP eigenstates of a time-evolving neutral B state that begins at time zero as B^0 to those of the state that begins as a $\bar{B}^0$:

$$a_{CP} = \frac{\Gamma(B^0(t) \rightarrow f) - \Gamma(\bar{B}^0(t) \rightarrow f)}{\Gamma(B^0(t) \rightarrow f) + \Gamma(\bar{B}^0(t) \rightarrow f)}. \quad (1.17)$$

The asymmetry can be expressed in terms of the CP parameter λ , with

$$\lambda = e^{-i\phi_M} \frac{\bar{A}}{A}. \quad (1.18)$$

Here ϕ_M refers to the phase due to the $B^0 - \bar{B}^0$ oscillation amplitudes. Within the SM the phase factor is given by [28]:

$$e^{-i\phi_M} = e^{-2i\beta + \mathcal{O}(10^{-3})}, \quad (1.19)$$

and β is the angle in the unitarity triangle. The A and $\bar{A}$ are the decay amplitudes for the $B^0 \rightarrow f$ and $\bar{B}^0 \rightarrow f$ decays, respectively, and $\bar{A} = \eta_f A$, with $\eta_f = \pm 1$ the CP eigenvalue of the final state f . The asymmetry is

$$a_{CP}(t) = \frac{(1 - |\lambda|^2) \cos(\Delta m t) - 2\mathcal{I}m(\lambda) \sin(\Delta m t)}{1 + |\lambda|^2}. \quad (1.20)$$

It is non-zero if any of the types of CP violation is present, $\lambda \neq \pm 1$. But, even if $|\bar{A}/A| = 1$ and $|e^{i\phi}| = 1$, with $|\lambda| = 1$ it is still possible that $\mathcal{I}m\lambda \neq 0$, which manifests purely the CP violation in the interference between mixing and decay. The asymmetry is typically re-written as

$$a_{CP}(t) = S_f \sin(\Delta m t) - C_f \cos(\Delta m t), \quad (1.21)$$

with

$$S_f = \frac{2\mathcal{I}m\lambda}{1 + |\lambda|^2} \quad \text{and} \quad C_f = \frac{1 - |\lambda|^2}{1 + |\lambda|^2}. \quad (1.22)$$

If amplitudes with only one common CKM phase dominate the decay, the $|A| = |\bar{A}|$ and $C_f = 0$, i.e., there is no direct CP violation, and $S_f = \eta_f \sin(2\phi)$, where η_f is the CP eigenvalue as before and 2ϕ is the phase difference between the $B^0 \rightarrow f$ and $B^0 \rightarrow \bar{B}^0 \rightarrow f$ decay paths. Without a contribution of another amplitude with a different CKM phase, the asymmetry is independent of strong interaction phases and therefore can be linked to fundamental CKM parameters with negligible corrections. In that respect, the $b \rightarrow c\bar{c}s$ decays to CP eigenstates such as $J/\psi K^0$ are theoretically cleanest, measuring $S_{J/\psi K^0} = -\eta_f \sin(2\beta)$. The $b \rightarrow s\bar{s}s$ penguin amplitudes dominating

the decay within the SM, $B^0 \rightarrow \phi K^0$, have, to a high precision, the same weak phase as the $b \rightarrow c\bar{c}s$ tree amplitudes. On the other hand, if there are additional amplitudes due to new generations of particles entering the penguin loops with different weak phases, one expects that

$$S_{\phi K^0} \neq S_{J/\psi K^0}, \quad (1.23)$$

and $C_{\phi K^0}$ could be non-zero. The expected difference between the two S_f in the SM is $\approx 4\%$ [5]. Hence, the measurement of asymmetries due to CP violation in the interference between mixing and decay are a clean way to detect the presence of Physics beyond the SM.

1.4 Decay Rate Measurement

The measurement of rates of charmless B meson decays provides input to models for the final state interaction. Strong interaction couplings cannot be derived from the Lagrangian of strong interaction, but the relation of rate measurements to fundamental CKM parameters depends on the precise knowledge of these couplings in the transition amplitudes. Several theoretical estimates for the branching fraction (BF) for $B^0 \rightarrow \phi K^0$, based on different assumptions, predict the range $(0.7\text{--}13) \times 10^{-6}$ [11].

First evidence for the decay $B \rightarrow \phi K$ has been seen by the CLEO experiment [31]. They measure in the charged mode:

$$BF(B^- \rightarrow \phi K^-) = (5.5^{+2.1}_{-1.8}(\text{stat}) \pm 0.6(\text{syst})) \times 10^{-6}. \quad (1.24)$$

For the neutral B decay mode $B^0 \rightarrow \phi K^0$, they report an upper limit:

$$12.3 \times 10^{-6} \quad \text{at 90\% C.L.} \quad (1.25)$$

This channel was first observed by *BABAR* [32]:

$$BF(B^0 \rightarrow \phi K^0) = (8.1^{+3.1}_{-2.5}(\text{stat}) \pm 0.8(\text{syst})) \times 10^{-6} \quad (1.26)$$

with a significance of greater than 5 standard deviations (statistical significance of 6.4σ).

Chapter 2

The *BABAR* Experiment

The data used in this analysis was recorded by the *BABAR* detector, a high energy physics experiment at the Stanford Linear Accelerator Center (SLAC) with the goal of studying the violation of charge and parity (CP) symmetry in the decays of B mesons. The experiment is designed and maintained by a collaboration of nearly 600 physicists from around the world. The idea for such a detector was proposed in 1987 by Piermaria Oddone [33]. He suggested that an asymmetric e^+e^- collider operating at the $\Upsilon(4S)$ resonance would be an excellent way to study CP violation in B decays. Eventually, two such experiments were created: Belle at KEK in Japan [34] and *BABAR* at SLAC in California [35]. *BABAR* was proposed in 1994, the PEP-II accelerator came online in 1998, and *BABAR* began taking data in May 1999. At the time of this analysis, *BABAR* has recorded over 200 million $B\bar{B}$ events and has published over 100 papers on CP violation and other physics topics.

2.1 PEP-II B Factory

The branching fractions of B decays to CP eigenstates are generally less than 10^{-4} and so a high luminosity is needed in order to produce a sufficient number of decays for study. In order to calculate the expected rate at which the *BABAR* detector records $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^0\bar{B}^0$ events we need to know three parameters: the PEP-II luminosity, the cross section of $e^+e^- \rightarrow \Upsilon(4S)$, and the decay rate of $\Upsilon(4S) \rightarrow B^0\bar{B}^0$. The luminosity, an expression of the “brightness” of the beam, is given in units of frequency per unit area. The design luminosity for PEP-II is $\mathcal{L} = 3 \times 10^{33} \text{cm}^{-2}\text{s}^{-1}$. The e^+e^- interaction occurs at the energy of the $\Upsilon(4S) = 10.58 \text{ GeV}$. Cross-sections for fermion pairs at the $\Upsilon(4S)$ are given in Table 2.1. We are interested in the cross-section $\sigma(e^+e^- \rightarrow b\bar{b}) = 1.05 \text{ nb}$.

Table 2.1: Production cross-section at the $\Upsilon(4S)$ for fermion pairs

$e^+e^- \rightarrow$	Cross-section (nb)
$b\bar{b}$	1.05
$c\bar{c}$	1.30
$s\bar{s}$	0.35
$u\bar{u}$	1.39
$d\bar{d}$	0.35

Finally, the $\Upsilon(4S)$ will decay to $B^0\bar{B}^0$ 47% of the time [23]. With these parameters, we calculate the $B^0\bar{B}^0$ production rate:

$$\Gamma_{e^+e^- \rightarrow B^0\bar{B}^0} = \mathcal{L} \cdot \sigma_{e^+e^- \rightarrow b\bar{b}} \cdot \sigma_{\Upsilon(4S) \rightarrow B^0\bar{B}^0} = 1.48 \text{ s}^{-1}. \quad (2.1)$$

The measurement of time-dependent CP asymmetries in the B meson system places several requirements on the design of the accelerator and detector. A high production rate can be achieved with electron-positron collisions at the $\Upsilon(4S)$ resonance energy, which result in production of a $B\bar{B}$ pair nearly 100% of the time. The $\Upsilon(4S)$ is only slightly more massive than a $B\bar{B}$ pair. When produced at rest each $B\bar{B}$ has a small momentum. This, coupled with the small B lifetime ($\tau_B = 1.5\text{ps}$), results in a very small spacial separation of the $B\bar{B}$ pair upon decay. Production of $\Upsilon(4S)$ with a boost (non-zero lab-frame momentum) increases the separation of B decay vertices, substantially improving the precision with which this separation can be measured. This motivates an asymmetry in the electron-positron beam energies.

The PEP-II facility at SLAC consists of two storage rings: the high energy ring (HER) for electrons with energy of 9 GeV and the low energy ring (LER) for positrons with energy of 3.1 GeV (shown in Figure 2.1). This energy-asymmetry produces a Lorentz boost to the $\Upsilon(4S)$ of $\beta\gamma = 0.56$. This boost increases the average B decay length of $80\mu\text{m}$ to $250\mu\text{m}$ in the laboratory frame: allowing measurement of the separation of the two B decay vertices with much higher significance. Additionally, the maximum B transverse momentum is $300\text{MeV}/c$: small enough that measurement of the z component of the B decay vertex is sufficient for determination of the relative B lifetime. The high interaction rate required by the experiment places a premium on maintaining a high beam current (number of electrons/positrons in the storage rings at any one time) and tight beam focus (cross-sectional area of the beam) since these factors directly relate to the $B^0\bar{B}^0$ production rate (Eq. 2.1).

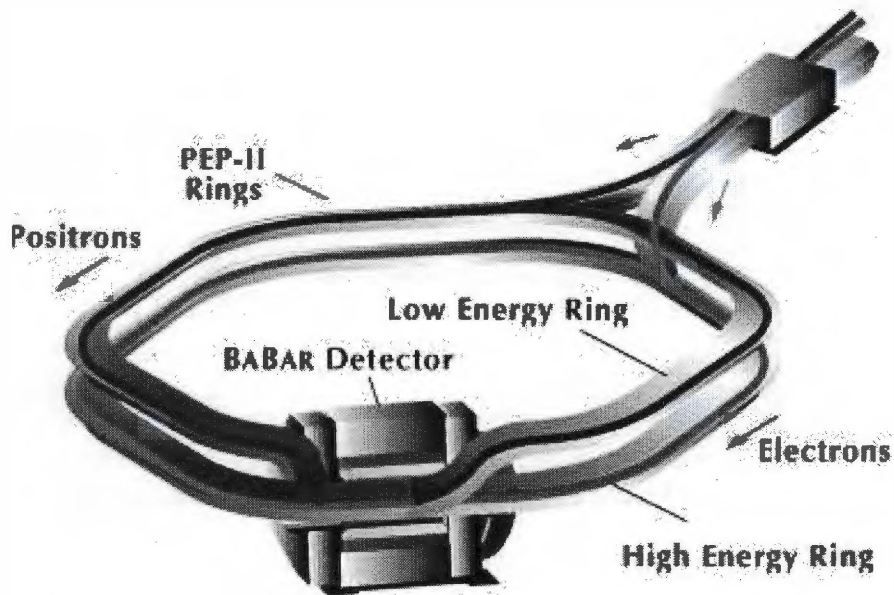


Figure 2.1: At the SLAC PEP-II accelerator facility, electrons and positrons are accelerated to high energies in the linear accelerator (not shown) and injected into two opposing storage rings.

While PEP-II operates at the $\Upsilon(4S)$ resonance most of the time, a limited amount ($\approx 12\%$) of data is taken 40 MeV below resonance for study of continuum background ($e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ events). This analysis uses 244 fb^{-1} of on-resonance data corresponding to 227 million $B\bar{B}$ pairs. Figure 2.2 shows the integrated luminosity both provided by PEP-II and recorded by *BABAR* since the experiment began.

2.2 The *BABAR* Detector

The *BABAR* detector is located at interaction region number two of PEP-II. The *BABAR* coordinate system puts the z-axis in the direction of the high energy beam and the y-axis directly up. Due to the asymmetric beam energies, the interaction point (IP) is offset from the center of the detector by about 0.37 m in the positive z direction. The detector sub-systems are located in a roughly cylindrical configuration around the IP.

A typical $B\bar{B}$ event contains an average of 11 charged decay particles. Full reconstruction of B meson decays, the typically small branching ratios of these decays, and the detector's other physics goals put many requirements on the performance and features of the detector [36]. Objectives include:

- Large uniform angular acceptance for charged and neutral particles down to small polar angles relative to the boost direction ($> 300 \text{ mrad}$),
- High reconstruction efficiency for charged particles with momenta as low as $60 \text{ MeV}/c$ and for photons with energies as low as 20 MeV ,
- Good momentum resolution for charged particle momenta above $60 \text{ MeV}/c$,
- Excellent energy and angular resolution for the detection of photons,
- Very good vertex resolution parallel to the beam direction,
- Efficient electron and muon identification with low hadron mis-identification,
- Efficient and accurate identification of hadrons (pions, kaons, and protons) over a wide range of momenta,
- Radiation hardened detector components (dose near the IP is 1 Mrad over 10 years),
- High bandwidth data acquisition and control systems,
- Detailed monitoring and automated calibration,

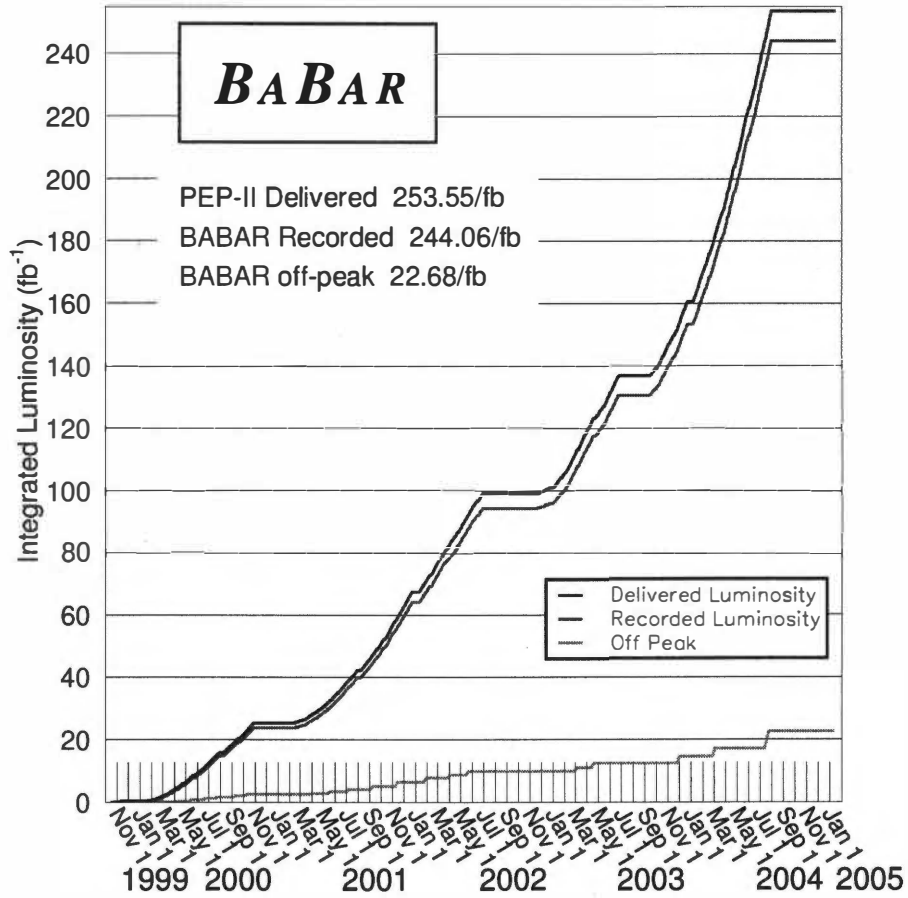


Figure 2.2: B Factory integrated luminosity since 1999 (start of physics data taking) until the end of this analysis showing the luminosity provided by PEP-II (darkest line), recorded by *BABAR* ($\approx 96\%$ efficiency) (middle line), and recorded at 40 MeV below the $\Upsilon(4S)$ resonance (lightest line)

- An online computing system capable of processing and simulating large numbers of events.

Figure 2.3 shows a three dimensional cutaway diagram of the detector and its major subsystems. Closest to the beam pipe is the Silicon Vertex Detector (SVT). The SVT employs the ionization of semi-conducting sensors by charged particles to measure their position and direction close to the IP. Progressing radially outward from the IP is the Drift Chamber (DCH). The DCH is used to measure the momentum of a charged particle as it moves through and ionizes the gas within the DCH. It also contributes specific ionization (dE/dx) measurements for particle identification (PID) purposes as well as tracking information. The next sub-detector in radial order from the IP uses a novel technique for particle species identification (PID) via the Detection of Internally Reflected Cherenkov Light (DIRC). The DIRC serves as the principal source for *BABAR* PID. The electromagnetic calorimeter (EMC) detects electromagnetic showers caused by neutral particles and high energy photons as they travel through a finely segmented array of Cesium Iodide crystals. Next outward from the beam pipe is the superconducting solenoid which provides the highly uniform 1.5 T magnetic field within the charged track detectors. Finally, the Instrumented Flux Return (IFR) detects muons and neutral hadrons via resistive plate chambers inserted between the steel plates that make up the magnetic flux return.

2.2.1 Charged Particle Tracking

The *BABAR* charged particle tracking system is designed to accurately and efficiently measure the tracks and momenta of charged particles (and their descendants) originating from the IP of the detector. Information from two sub-detectors, the SVT and DCH, are combined in order to parameterize each particle track. The particle momentum is calculated with information from the track's curvature in the 1.5 T magnetic field running parallel to the z-axis of the detector.

Silicon Vertex Tracker

The SVT is a five layer detector composed of double sided silicon strip sensors. Figure 2.4 shows the three innermost layers starting from 3.2 mm from the beam pipe and the two outer layers at 144 mm from the beam pipe and above. The close proximity to the IP enables the SVT to very accurately measure particle decay vertices (as are required in *B* decay-time measurements) and to reconstruct tracks of low p_T particles that would otherwise escape detection.

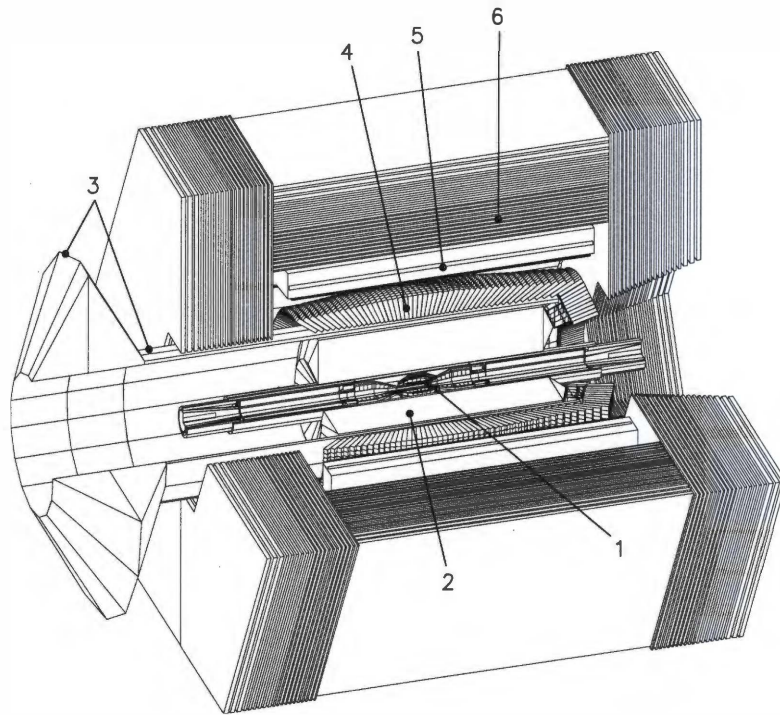


Figure 2.3: Cutaway diagram of the *BABAR* detector that shows the Silicon Vertex Tracker (1), Drift Chamber (2), Cherenkov Detector (3), Electromagnetic Calorimeter (4), superconducting solenoid providing 1.5 Tesla magnetic field (5), and Instrumented Flux Return (6)

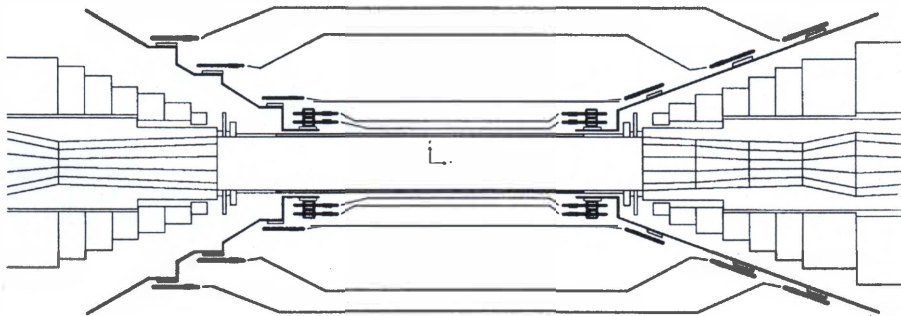


Figure 2.4: Details of the inner parts of the detector, showing the cross-sectional view of the silicon vertex tracker (SVT) in a plane containing the beam axis

The silicon sensors detect the passage of a charged particle via the production of electron-hole pairs within the semiconductor. A potential difference across the sensor causes the electrons/holes to drift to contacts on opposite sides of the silicon sensor. The charges are collected (registered) by an array of parallel metal strips on either side, which are rotated by 90° with respect to each other. This arrangement allows reconstruction of the hit in both the z and ϕ directions.

The three inner layers are positioned for optimal vertex measurement, while the outer two layers are near the DCH in order to facilitate corroboration of tracking information between the DCH and SVT. The silicon wafers are positioned in an overlapping pattern in order to maintain high acceptance, and the outer layers are curved toward the beam pipe in an effort to maximize the coverage while conserving the amount of silicon used. Due to the proximity to the beam pipe, the SVT must be particularly resistant to radiation and is equipped with an array of radiation monitoring photo-diodes connected to a beam shutdown circuit.

The SVT has a minimum resolution of $20\text{ }\mu\text{m}$ in ϕ and $40\text{ }\mu\text{m}$ in z for tracks traversing the sensor at a 90° angle.

Drift Chamber Tracking

The DCH measures the track coordinates, momenta, and specific ionization of charged particles. It consists of a 2.8 m long cylindrical chamber surrounding the SVT (between a radius of 23.6 cm and 80.9 cm) within the 1.5 T magnetic field of the detector. It provides an average of 40 tracking hits per charged track for particles having transverse momentum greater than $100\text{ MeV}/c$.

The DCH reconstructs particle tracks by detecting the secondary charges liberated as the charged particle traverses the DCH gas medium (80% helium, 20% isobutane). The ionization creates electrons and ions within the gas, which drift in an electric field shaped by field wires strung along the z -axis of the detector. The electrons accelerate in the electric field created by the high-voltage wires ($\approx 1930\text{ V}$). They gain enough energy to create further ionization along their path. This generates an avalanche of charged particles, which can then be detected by the grounded, or “sense,” wires. Average gain due to this avalanche effect is about 5×10^4 .

The DCH uses 7,104 sensing wires each surrounded by 6 high voltage wires (forming one drift cell). The DCH consists of 40 layers of cells in 10 groups (Figure 2.5). Some groups of cells are strung purely in the z direction while others are given small (alternating) positive or negative stereo angles in order to facilitate measurement of the track’s

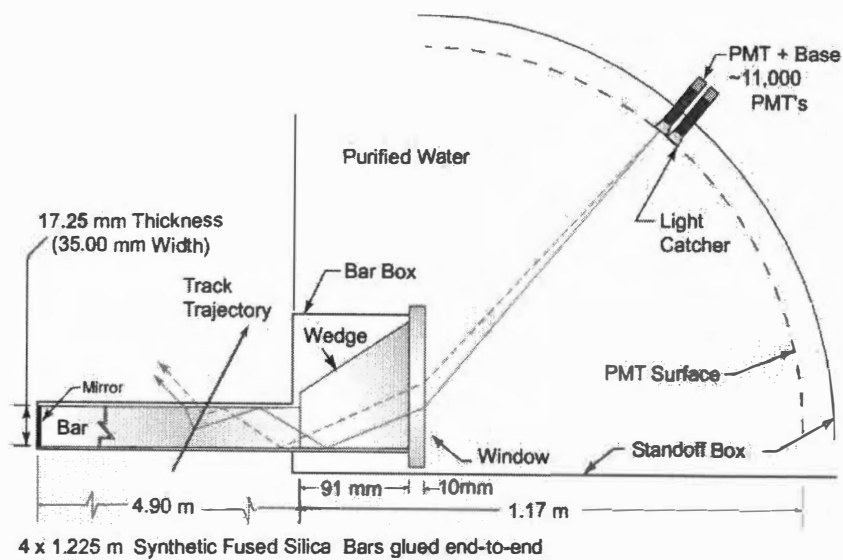


Figure 2.6: This schematic view of one DIRC quartz bar shows how the Cherenkov light cone is reflected within the quartz bar and guided to the readout PMTs within the water-filled standoff box.

DIRC that illustrates the principles of light production, transport, and imaging. The radiator material of the DIRC is synthetic fused silica in the form of long thin bars (surrounded by nitrogen) with rectangular cross section of $3.5 \times 1.7 \text{ cm}^2$. The bars serve both as radiators and as light guides for the Cherenkov light. For high velocity particles traversing the DIRC, Cherenkov light is emitted under a characteristic angle θ_c with respect to the particle trajectory according to the relation $\cos \theta_c = 1/n\beta$. Here, θ_c is the Cherenkov angle, the velocity $\beta = v/c$ (c = the speed of light in the medium), and $n = 1.473$ is the mean index of refraction for silica. The particle's mass (and thus particle species) can be determined from the measurement of the Cherenkov opening angle (θ_c) and momentum information from the tracking system. Figure 2.7 shows the correlation between Cherenkov angle and the laboratory momentum of the decayed particle.

For particles with $\beta \approx 1$, some photons will always lie within the total internal reflection limit and be transferred to either end of the bar. A mirror is placed at the forward end of the bar in order to reflect photons back toward the rear, instrumented, end of the DIRC. This rear side of each silica bar is connected to a water-filled expansion region: the stand off box (SOB). The index of refraction of the water, $n = 1.33$, prevents total internal reflection at the quartz-water interface and allows the photons to enter the water volume. Each 12 bars are placed in an aluminum box, and 12 bar boxes are placed around the DCH at a minimum radius of 80 cm. The photons are detected by an array of $\sim 11,000$ photo-multiplier tubes (PMTs) arranged along the internal face of the toroidal stand off box. A silica wedge at the end of each bar reflects the photons at large angles relative to the bar axis, into the center region of the SOB, thus reducing the size of the required detection surface.

The DIRC occupies 80 mm of radial space in the central detector volume with a total of about 17% radiation length thickness at normal incidence. The distance from the end of the bar to the PMTs is $\sim 1.17 \text{ m}$. This, together with the cross-sectional size of the bars and the PMTs, contributes to the single photon Cherenkov angle resolution of $\sim 7 \text{ mrad}$. The uncertainty due to the unresolved wavelength dependence of the refractive index is $\sim 6 \text{ mrad}$. The overall single photon resolution is estimated to be about 10 mrad.

The average resolution of the Cherenkov angle measurement is 2.5 mrad. The DIRC kaon/pion separation per track is about 4σ at a particle momentum of $3 \text{ GeV}/c$. Beam background photon suppression is achieved with precise measurements of the photon arrival time (1.5 nsec resolution per track). The highest background hit rate per PMT is 200 kHz. The error on this value is dominated by the number of photons counted (average of 28 for a normally incident $\beta = 1$ particle) and the track direction uncertainty.

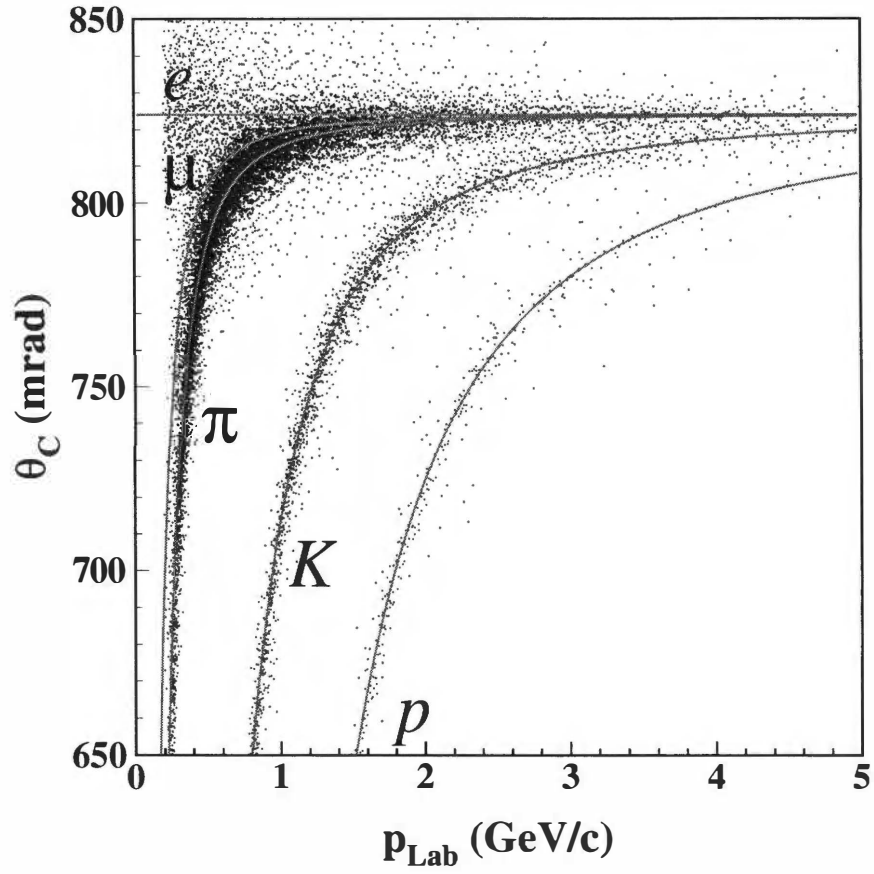


Figure 2.7: The measured Cherenkov angle of tracks from an inclusive sample of multi-hadron events plotted against the momentum of the tracks with gray lines indicating the predicted values of θ_C for different particle species.

DCH dE/dx Measurement

For lower momentum particles of insufficient speed to be detected by the DIRC, PID is still possible via measurement of the specific energy loss (dE/dx) as the particle traverses the tracking system. The total energy loss in the DCH is calculated from the total charge deposited on the sense wires during the particle's trip across the drift chamber cells. The correlation with the particle momentum [37] permits determination of the particle species for low momentum tracks (see Figure 2.8).

2.2.3 Electromagnetic Calorimetry

Of particular importance for this analysis is the measurement of the energy and location of the four photons from the neutral kaon decay ($K_s^0 \rightarrow \pi^0 \pi^0$, $\pi^0 \rightarrow \gamma\gamma$) with the Electromagnetic Calorimeter (EMC): consisting of 6580 thallium doped cesium iodide (CsI(Tl)) crystals.

Photon Detection

When a high energy photon (required energy greater than 20 MeV) enters one of the CsI(Tl) crystals, it can interact in the strong electric field of one of the atoms and create an electron-positron pair. Both of these particles can then emit photons via bremsstrahlung as they accelerate in the electric field of atoms in the material. With photon energies high enough, the process continues as the newly created high energy photons themselves initiate pair production. Thus, the number of particles in the direction of the primary photon (sharing the initial photon's energy) increases in a process known as an electromagnetic *shower*. Eventually, the shower begins to decline as the average energy of the produced particles falls below a threshold E_c and other processes become dominant (Table 2.2). The longitudinal shower length is proportional to the energy of the initial photon. The crystal length is chosen to contain the shower of a 2 GeV photon. The lateral spread of the particles in the shower is affected by Coulomb scattering and increases with the shower depth. The lateral size is measured in terms of the Molière radius (R_m). By definition, 95% of the shower energy is contained within one R_m . The R_m should be kept small since it limits the angular resolution of the EMC.

Basic interaction processes of lower energetic photons in material are listed in Table 2.2 together with their dependence on the energy of the primary photon. In all three processes, the secondary charged products that carry the energy information are electrons and positrons. In CsI crystals, these charged particles transfer kinetic energy to the lattice by exciting electrons to higher energy levels and creating electron-hole pairs.

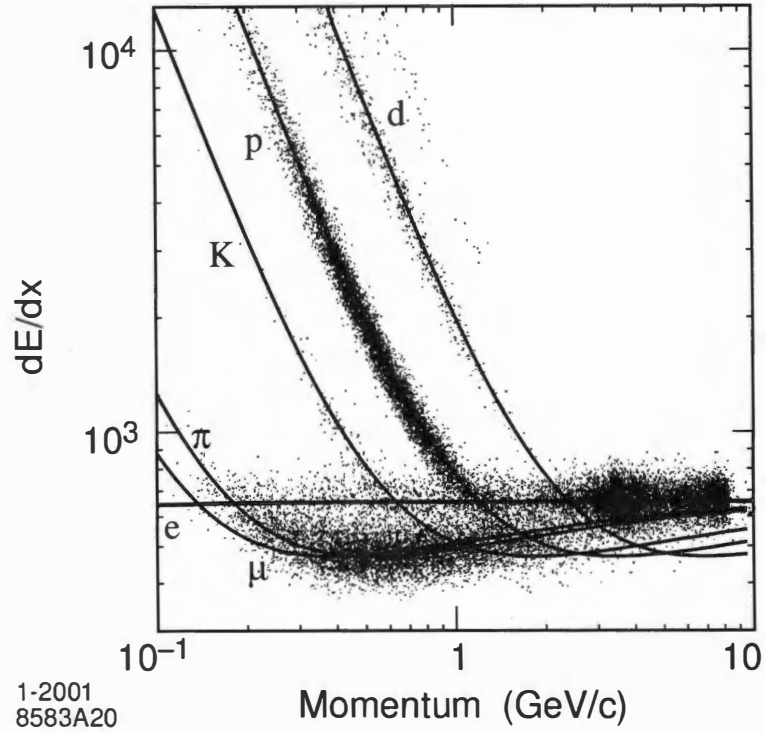


Figure 2.8: dE/dx information from the drift chamber plotted as a function of particle momentum shows clear differentiation between particle species.

Table 2.2: Interaction processes of photons with material in different regimes of the photon energy E

Process	Photon energy domain
Photo-effect	$E < 0.1 \text{ MeV}$
Compton scattering	$0.1 < E < 3 \text{ MeV}$
Pair production	$E > 3 \text{ MeV}$

Excited Cs and I atoms emit the energy in form of electromagnetic radiation in the UV region. The recombination of electron-hole pairs may lead to emission of light when this excites atoms of the lattice. This UV light is easily reabsorbed and is difficult to detect. Electron-hole bound states (excitons) moving through the lattice also can transfer their binding energy to activator centers (impurities like the thallium atoms) which subsequently de-excite. These activator centers have more closely spaced energy levels and emit light in the visible region. The number of *scintillation* photons depends on the type of ionizing particle, here only electrons, and the conversion is nearly linear over a wide energy range. In thallium doped CsI crystals, about 20 eV are required to produce one scintillation photon.

EMC Geometry

The EMC crystals are arranged in two configurations within the detector (see Figure 2.9). The spacial distribution of crystals is sufficient for about 90% acceptance of photons emitted in the center of mass frame. The majority of crystals in the EMC are placed at roughly constant radial distance (about 92 cm inner radius) in a barrel configuration with crystal faces pointing toward the IP. The remaining crystals are placed in a forward direction end cap to account for Lorentz-boosted particle trajectories. Each crystal has a front face of about $47 \times 47 \text{ mm}^2$ (approximately equivalent to one Molière radius in CsI), a rear face of about $60 \times 60 \text{ mm}^2$, and an average length of 30 cm. The scintillation light is detected with a pair of silicon photo-diodes attached to the rear surface of each crystal. These photo-diodes are optimized for sensitivity at the average wavelength of the scintillation light and have a detection efficiency of 85%. An average of 7,300 photons are detected per MeV of primary photon energy.

The energy measurement is dependent on an absolute calibration. This is achieved with a liquid radioactive photon source circulated in front of the crystals and by kinematic modeling. The energy resolution for photons at a 90° polar angle is

$$\frac{\sigma_E}{E} = \frac{1\%}{\sqrt[4]{E(\text{GeV})}} \oplus 1.2\% . \quad (2.2)$$

The constant term arises from front and rear energy leakage, mis-calibration, and light collection non-uniformity. The angular resolution ranges between 12 mrad at low energies to 3 mrad at high energies.

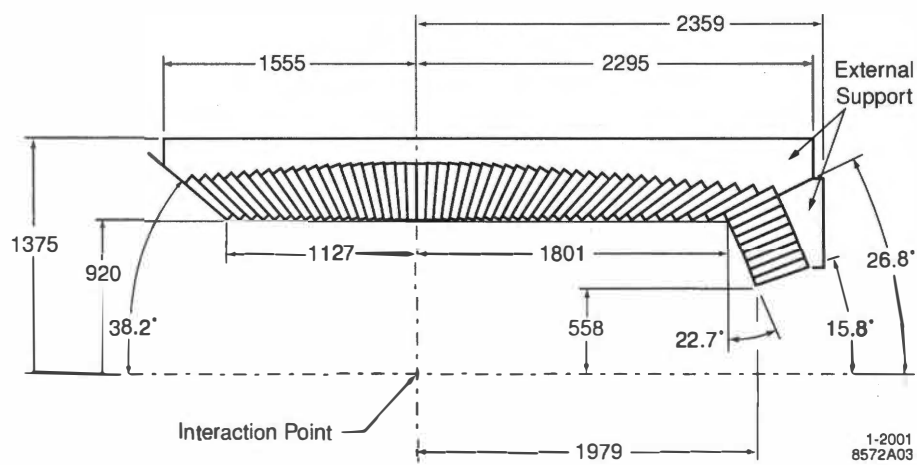


Figure 2.9: A side view of one line of EMC sensors showing both the barrel section and forward end-cap (to account for the boosted center-of-mass frame) with units in mm

Table 2.3: Overview of the angular coverage (in the *BABAR* center-of-mass frame) for each detector sub-system

Sub-detector	Solid angle coverage in cm frame ($/4\pi$)
SVT	0.90
DCH	0.92
DIRC	0.87
EMC	0.90 (0.92 @ 100 MeV)
IFR	0.94

2.2.4 Instrumented Flux Return

The Instrumented Flux Return achieves muon and neutral hadron identification via a system of resistive plate counters located outside the radius of the EMC. Information from the IFR was not used in this analysis.

2.2.5 Summary

Summary of solid angle acceptance and measurement resolutions for the *BABAR* detector subsystems are presented in Table 2.3.

Chapter 3

BABAR Computing Infrastructure

3.1 *BABAR* Monte-Carlo Simulation Production

This analysis makes use of the *BABAR* simulation production effort in order to estimate the systematic error related to detector understanding and in characterization of particle decays of interest. We require an in-depth understanding of detector-particle interaction, detector response to specific particle decays, and detector behavior in various environmental and operational conditions. The simulation also aids in the development of models used to characterize previously unseen particle decays and to identify and suppress sources of background. To address these needs we must integrate our simulated observations over all possible particle decays and particle trajectories for signal and background. To deal with the very large size of the variable space we use a statistical sampling method of integration known as the Monte-Carlo [38] method.

3.1.1 The Monte-Carlo Method

The Monte-Carlo (MC) method is a numerical integration method that is particularly well suited for complex scientific problems that have a large number of parameters and therefore a large solution space. These types of problems can be impossible to characterize using direct numerical integration techniques. By instead taking a random sample of states from the solution space of the problem, we are able to drastically reduce the computing time required in order to get an answer of practical significance (low enough error). The MC method yields an answer with an error proportional to $n^{-1/2}$, where n is the number of samples taken. Most simple numerical integration methods have an error proportional to n^{-2} , so it would appear that MC is a much less efficient technique. However, for multi-dimensional problems where numerical integration error is proportional to $n^{-2/d}$ (d is the number of dimensions), the error for MC integration remains

proportional to $n^{-1/2}$. Therefore, the MC approach is advantageous for any problem with five or more free parameters.

3.1.2 The *BABAR* MC Simulation

The *BABAR* collaboration uses a MC simulation based on the GEANT 4 [39] detector simulation package from CERN. The simulation involves three steps: event generation, detector model, and event reconstruction.

Event Generation

The first step is for the event generator to choose the decay tree and reaction product kinematics. The generator takes commands from a decay configure file which constrains the decay tree to a varying degree. Starting with the initial e^+e^- interaction, we can constrain the chain of decays to the desired sub-tree of the full e^+e^- decay space. For instance, in $B^0\bar{B}^0$ generic MC, the decay is constrained to follow the chain $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^0(\bar{B}^0)$: it is free to explore only the decay tree for B^0 and $\bar{B}^0$. Further decays are randomly chosen based on known branching fractions from a look-up table. For signal MC, one B^0 ($\bar{B}^0$) will decay to a characteristic final state, and the second B^0 ($\bar{B}^0$) will follow a randomly chosen decay. In the case of B^0 and $\bar{B}^0$, all time-dependencies (decay, mixing, CP violation) are modeled as well.

Detector Model

Once the initial decay tree is generated, it is then placed within a simulation of the *BABAR* detector. Particles that live long enough to reach detector material are tracked in small volume intervals. In each volume, interaction of the particles with the detector material and environment is evaluated. Based on the survival probability, (from PDG tables [23]) the particles are decayed and its decay products are propagated further. Secondary particles from material interactions (such as liberated electrons) are propagated as long as they are above certain momentum or energy thresholds or until they leave the detector volume. These secondaries indicate the path of a particle's trajectory if registered in sensitive volumes of the detector. Since the detection method is modeled as well, from this stage on the simulated event appears like a raw data event and is processed analogously.

The simulation includes information on the detector materials, the configuration of the detector at a given time (such as DCH voltages and magnetic field strength), and the conditions of the detector environment (like humidity and temperature). At

this stage, real beam-background events are mixed into the simulated event from data lookup tables. Both detector configuration data and backgrounds are selected in order to accurately represent a particular time in the detector’s data-taking history.

Event Reconstruction

The event reconstruction uses the information about the event’s interaction with the detector’s sensors and tries to interpret the nature of the event. It uses the same algorithms that are applied to real data events to reconstruct the event from the detector’s sensor readings.

Also added to the MC data is information about the “true” decay that was produced from the event generator. This allows calculation of detector efficiencies, resolutions, and detailed studies of other detector responses.

3.1.3 Use of MC in *BABAR* Analysis

The simulation of *BABAR* data is a computing challenge compounded by many factors. The large variety of physics processes accessible to the detector, changing configuration and backgrounds in the detector and accelerator, and changes in software due to improved physics understanding place large demands on the available computing resources.

In order to reduce the statistical uncertainty in quantities that are dependent on the amount of reconstructed MC events, we must produce MC for many times the number of recorded events. The *BABAR* collaboration approaches this need with computing resources at many different sites around the world. We will present details about the contribution of the High Energy Physics Group at The University of Tennessee.

3.1.4 The Simulation Production Effort

The Simulation Production (SP) of *BABAR* relies on computing sites around the world. These sites range from large national computing centers like SLAC to smaller efforts operated by universities. Using a standardized set of simulation jobs, the different hardware and software components and SP databases are validated. Each set consists of three standard decay trees for each month of data taking. Each period of data taking produces a background collection and corresponding conditions collection which are used together in MC simulations for a given data taking period. The MC tests span the full range of background and conditions collections that will be used for MC production. The tests are divided into “runs” of 2000 simulated events with each run corresponding to a

single combination of decay tree, background collection, and conditions collection. Reconstructed quantities such as certain particle multiplicities are collected into histograms and compared to a standard set provided by the reference site, SLAC.

The scheduling of jobs (SP MC requests), update of software releases, and maintenance of SP databases are coordinated between the local site manager and the overall SP coordinator. We will describe the methods and solutions which we developed in the course of our SP production effort for *BABAR* and were provided as tools to other sites.

3.2 Computing System Design

The computing system is specifically designed for the purpose of producing Monte-Carlo simulated events with the *BABAR* MC simulation production software and management utilities. The MC simulation is allocated in increments of no more than 2000 events per job. Each job takes between 1 and 6 hours to produce 2000 simulated events. The simulation executable requires a minimum of 500MB of RAM while running to avoid inefficiencies caused by use of disk swap (virtual memory). Each job produces a 35MB MC data collection which needs to be stored locally until it can be transferred to the central MC database at SLAC. This means that a site capable of running 80 simultaneous simulation jobs would require about 20GB of disk storage to stage a full day of MC simulation data. Since the MC production software can be easily run on multiple machines simultaneously, it is a perfect candidate for a batch-queue compute cluster type system.

This system usually consists of a large number of powerful computers (“compute nodes”) which are connected via a switched network to one or more central servers (see Figure 3.1). The central servers send compute requests to the compute nodes and then receive and organize the results when they are produced. This type of system makes efficient use of CPU resources and is highly scalable because the individual computers that make up the cluster can work almost independently of one another.

When designing a computation infrastructure there are competing factors such as performance, cost, maintenance, and construction time. The compute node hardware has to maximize the amount of aggregate computing power available for MC production. We only included sufficient support hardware in the design so as to ensure efficient use of the compute node CPU resources. Components of the system which do not directly contribute to the computation effort are eliminated from the design. Our design seeks to minimize maintenance time by including redundancies where failures are most likely to occur and by reducing or eliminating unnecessary components that are prone to failure.

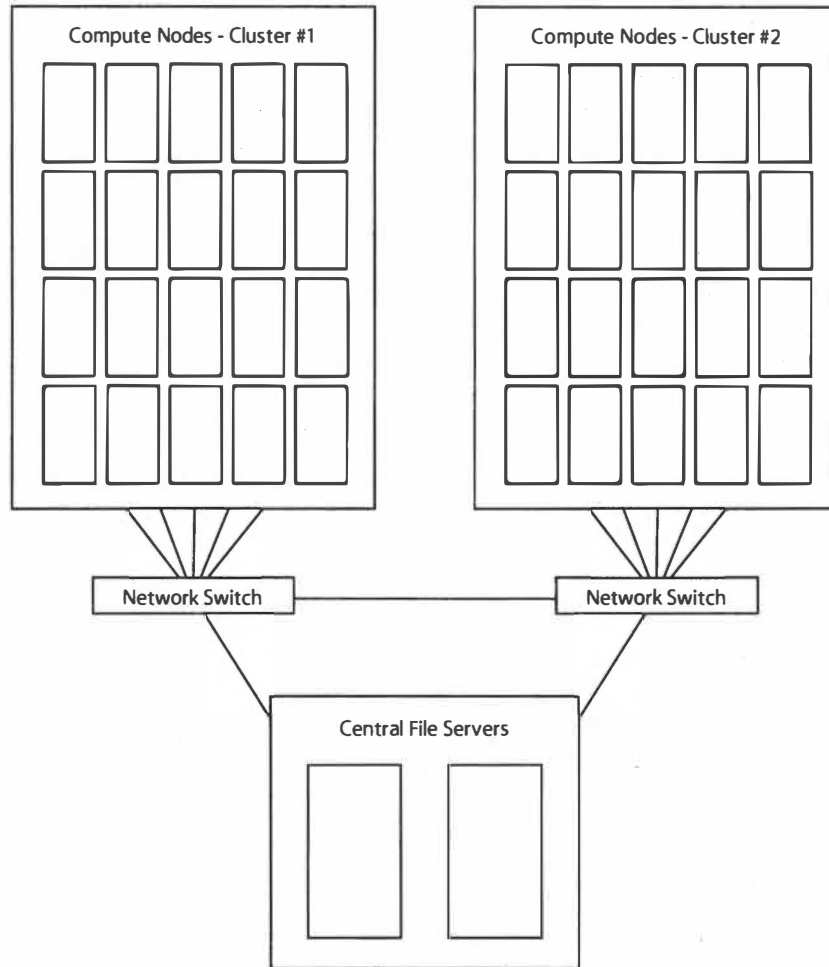


Figure 3.1: This is a schematic of the compute cluster configuration with the two central file servers connected to the network switches via 1 Gbit/sec uplink ports. The switches are connected together and then connected to the compute nodes via 100 Mbit/sec ports. The compute nodes are separated into two groups, each connected to one of the network switches.

Finally, the compute node enclosure hardware was selected with a specific interest in heat management since poor heat dissipation is a major cause of hardware failure in computer systems.

3.2.1 Diskless Compute Cluster

The most innovative aspect of the compute cluster design is that the compute nodes operate in a diskless configuration. The nodes have no hard disk drive or other form of permanent storage installed. Instead they receive everything that they need (operating system files and MC simulation software) from the central file servers. This design offers economic and administration advantages over a conventional disk-based compute node.

Upon startup, each compute nodes transfers control of the boot process to a small computer program executed off a read only memory (ROM) chip on the network card. This software, called the Managed Boot Agent (MBA), initializes the network card and attempts to contact a remote computer to get an internet protocol (IP) address. A program running on the central file server, called the Dynamic Host Configuration Protocol Daemon (DHCPD), sees this IP address request and responds with an assigned IP address from a lookup table maintained on the server. This response also includes a network location where the compute node can find a specially prepared Linux kernel file (the core of the Linux operating system). The compute node then downloads the Linux kernel from this network location, executes it, and transfers control of the boot process to the Linux kernel. The kernel then continues the boot process by initializing the rest of the compute node hardware. The file systems used by the compute node (including the root file system) are located on the central file servers and are mounted remotely via the Network File System (NFS) protocol.

Many diskless compute clusters of this type create a separate file system on the central servers for each compute node in the cluster. Since the file system used by the compute node can be rather large in size (around 10 GB), a lot of disk space is required. Additionally, software updates and configuration changes become more complex because multiple file systems need to be changed in parallel. Since all compute nodes in the cluster are to be configured identically we avoid redundantly stored files for each node by configuring them to use a single system image for their root file systems.

The single compute node file system (located on the file server) is mounted by all 40 compute nodes. Conflicts and file corruption can occur when multiple compute nodes attempt concurrent writes to the same file. We solve this problem by only allowing the compute nodes to mount the file system in read-only mode. For directories that

require write access, we mount the directories as a memory-resident file system using the Random Access Memory (RAM) on each individual compute node. Since this file system is resident in volatile memory, a reboot or power failure will cause all files there to be lost. This is not a problem since these file systems only contain temporary state information that is not needed after a reboot. Upon reboot, the compute node configures these RAM file systems with default values from a known-good configuration.

A diskless configuration has many advantages over traditional disk-based compute nodes. By eliminating the hard disk drive, the compute nodes cost is reduced. Since the disk drive is one of the most failure prone components, this also increased the overall hardware reliability. Software configuration of the compute nodes is simplified by using a single system image for all compute nodes. Any software change made in one location on the central server automatically propagates to all the compute nodes. Finally, by using a read-only single system image, the software configuration is guaranteed to be stable and consistent among all the compute nodes in the cluster.

3.2.2 Compute Node Configuration

The compute node hardware was selected to maximize the ratio of MC events produced per dollar spent. The hardware for each node consists of two central processing units (with heat sinks and fans), a motherboard, RAM modules, a network interface card, an AC-DC power supply, and an enclosure with intake and exhaust fans. Figure 3.2 shows the internal hardware configuration of one of the compute nodes.

Central Processing Unit (CPU)

CPU performance was benchmarked by running the *BABAR* MC simulation software on a range of CPU types from Intel¹ and AMD.² We then extrapolated the performance linearly with CPU clock frequency within each processor family. The current CPU prices were obtained from competitive resale channels. We chose to base the compute node hardware around the Athlon MP 2000+ CPU from AMD since it was both the most powerful processor available and was significantly cheaper than comparable products from other manufacturers. The motherboards for these processors were also cheaper than similar motherboards for the other CPU manufacturers.

Another important factor in the choice of this CPU is that it supports Symmetric Multi-Processor (SMP) computing. This allows use of two CPU chips with a single

¹Intel Corporation, <http://www.intel.com/>

²Advanced Micro Devices, Inc., <http://www.amd.com/>

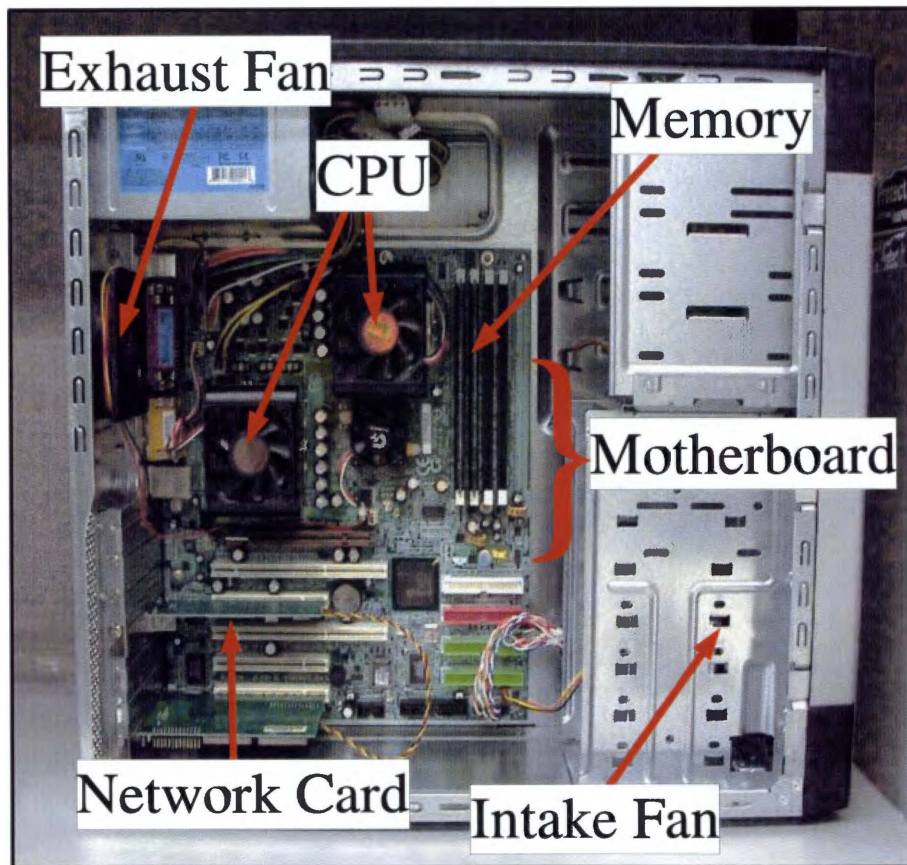


Figure 3.2: The compute nodes only contain components essential for operation: the motherboard, CPUs, RAM, network card, two fans, and a PC power supply.

motherboard. This saves space and hardware funds by sharing the case, power supply, and motherboard between the two CPUs. However, this economy of scale does not continue to higher number of CPUs because the motherboards and processors needed to run in groups of 4 (and higher) are much more expensive.

Heat production of the CPU is another important factor. We found that using the manufacturer recommended CPU fan and heat sink provided sufficient heat transfer to ensure reliable operation.

Network Interface Card (NIC)

In order to enable network boot and diskless compute node operation, the network card for each compute node has a special ROM chip. This ROM chip contains a program called PXE [40] (Pre-Execution Environment), which takes over control of the computer's boot process and downloads the Linux kernel from a computer on the network. We chose the 3Com 3C905CXi-TX-M³ 100 Mbit/sec network card with support for PXE for network booting. This card supports simultaneous transmission and reception (full duplex) of TCP/IP protocol [41] Ethernet. By taking the estimated MC job length (4 hours) and total network throughput during this time (100 MB) we determined that an average throughput of about 10 Mbit/sec was needed. Therefore, the 100 Mbit/sec provided by this NIC is sufficient.

This network card also has driver support in the Linux kernel. This is important because we need to compile the network card driver into the kernel in order to boot over the network.

Motherboard

We required that the motherboard provide support for two AMD Athlon MP 2000+ CPUs, 1 GB of RAM, and a PCI (Programmable Computer Interface) slot for the NIC. The only relevant performance characteristic of the motherboard is the front-side bus speed (FSB) (the speed in clock cycles per second at which the CPU can address the system memory). The required maximum FSB clock speed is determined by the CPU model (400 MHz for the AMD Athlon MP 2000+). Accordingly, two motherboards were chosen: the ASUS⁴ model A7M266-D and Gigabyte⁵ model GA-7DPXDW+.

³<http://support.3com.com/infodeli/tools/nic/3c90x/ethercd/english/usrguid.pdf>

⁴Asustek Computer Inc. <http://asus.com/>

⁵Gigabyte Technology Co., Ltd. <http://www.gigabyte.com.tw/>

Random Access Memory

The *BABAR* simulation production software requires at least 512 MB of RAM per running executable. Since each compute node has two CPUs, we run 2 executables on each node and, therefore, need at least 1 GB of RAM per node. Each compute node has 1 GB of unregistered PC2100 SDRAM type RAM.

Enclosure

The compute node enclosure (case) serves as hardware mount and promotes air flow for cooling. We use cheap and easily available desktop PC cases. These cases have large ports in the front and back equipped with fans rated to a minimum of 30 cubic feet per minute. The cases are stacked backward on metal shelves (see Figure 3.3) to allow easy access to the power and network connections for each compute node.

Other hardware

The advantages of building the compute nodes in-house include the ability to exclude components that we do not need. The compute nodes have no video cards, sound cards, monitors, speakers, or other components that are not directly required for operation of the simulation software.

3.2.3 Central Servers

The central servers for the cluster provide permanent (disk) storage and host the control and administration programs for the compute cluster. CPU-intensive applications are rarely executed on these servers, so their configuration has been optimized for maximum storage capacity and reliability.

All storage is protected from individual disk failure via RAID [42] (Redundant Array of Inexpensive Disks) hard disk drive systems. We operate two file servers, each with a 12-port 3ware⁶ Escalade RAID controller card. The RAID card interfaces up to 12 individual hard disk drives with the PCI bus on the motherboard. We use a combination of RAID levels depending on the use patterns of the file systems.

RAID level 1 is a simple mirroring system that pairs disk drives and writes identical data to each drive, providing redundancy and improved read performance with a 50% usage efficiency of the raw storage capacity of the disk drives. RAID level 5 groups a

⁶Applied Micro Circuits Corp. <http://www.3ware.com/>



Figure 3.3: The full rack of 40 compute nodes that make up the 80 CPU cluster consist of machines based on Athlon MP 2000+ (left) and Athlon MP 2400+ (right) CPUs. Two network switches are located above the second level of nodes. The nodes are mounted facing the rear of the rack in order to facilitate access to the network and power connections and to exhaust waste heat into the center of the room.

minimum of 3 disk drives while using a system of distributed parity⁷ data for redundancy. This provides for failsafe against single-drive failure as well as improved read performance with the cost of diminished write performance. Raw storage efficiency ranges between 66% and 92% depending on the number of disks used. RAID level 10 uses a minimum of 4 disk drives by pairing drives into RAID 1 groups and then non-redundantly distributing data storage evenly between groups. This provides improved read and write throughput as well as protection against the failure of a maximum of 50% of the disk drives. Raw storage efficiency is 50%.

We use RAID 5 for file systems requiring high capacity with low access speed requirements (such as for long-term data storage) and RAID 10 for file systems requiring high read/write performance (such as those used for data output from the compute cluster nodes). We have found that an array using eight 7200 RPM ATA disk drives in a RAID 10 configuration provides sufficient random read/write performance for simultaneous access from all 40 compute nodes. The servers currently have a combined storage capacity of nearly 4 TB.

Disk drives use the +12 V output from the power supply at a typical maximum operating current of 3 A. This mandates a power supply with a +12 V current rating of at least 36 A. The power supplies on these servers provide power for the 12 hard disk drives via multiple independent +12 V circuits. Start-up power requirements of the disk drives are higher than the typical operating levels. However, the RAID controller card staggers power-up of the disks to avoid a large increase in power use during start-up.

3.2.4 Network

There is a negligible amount of communication between the compute nodes. Most network capacity is needed for the connections to the central servers and between switches. Therefore, a system that increases bandwidth in these areas was chosen. The cluster uses a pair of 24 port Dlink⁸ 100 Mbit/sec switches to connect the compute nodes and central file servers. Each compute node is connected to a 100 Mbit/sec port via Category 5 [43] network cable. The switches are connected to each other and to each file server by a 1 Gbit/sec uplink port.

During normal MC production, a typical job transfers about 35 MB of data to the file server over a period of four hours. When running 80 jobs in parallel, the expected average throughput is about 1.5 Mbit/sec. A worse-case scenario is 80 jobs that individually take

⁷A parity bit is a binary digit that indicates whether the number of bits with value of one in a given set of bits is even or odd. This allows the detection and repair of any single bit-flip.

⁸D-Link Corporation/D-Link Systems, Inc. <http://www.dlink.com/>

1 hour to run: an average throughput of 6 Mbit/sec. This is well within the operating performance of the hardware components used in the cluster.

3.2.5 Electrical Power

To ensure adequate power availability and safety, we had sufficient dedicated 120 V circuits installed to support a compute cluster of up to 80 compute nodes and several file servers. A power interruption is most potentially damaging for computer systems when writing files to storage. This can cause file system corruption and loss of data. Although the file servers use journaling file systems to prevent file system damage, we took the added precaution of installing uninterruptible power supplies (UPS) on each of the file servers. These UPS units are capable of powering the files servers for a maximum of 30 minutes. This time is sufficient to ensure that most power surges and short blackouts will not affect the machine. In the event of an extended power outage, the computer will have time to finish writing files to disk and then shut itself off. The UPS battery status is monitored by a process on the server via the RS232 port. The UPS units also contain power conditioners to protect the file servers from voltage surges.

UPS power protection is not provided to the compute nodes. The compute nodes do not contain disk drives that could be affected by a power outage. However, each compute node is protected from voltage spikes by an inexpensive surge-protecting power distribution strip. Should the compute nodes lose power, all jobs that the machines are currently running will abruptly stop and be marked as “failed” by the MC production software. The failed jobs can then be easily restarted with the loss of an average of only two hours of computing time.

3.2.6 Air Conditioning System

By measuring the power use of one test machine under full load we estimate that the total power use of our cluster of 40 dual processor machines is 25kW. Assuming that all the electrical energy is converted to heat, this translates into 7.1 tons of refrigerating capacity or 8.9×10^7 Joules/hour. This required an upgrade to the air conditioning system.

3.3 Software Environment

3.3.1 Operating System

All machines in the compute cluster run a distribution of the GNU/Linux operating system called Scientific Linux⁹ version 3.03. This Linux distribution is a binary-compatible re-distribution of Red Hat Enterprise Linux¹⁰ version 3.03. It was started by a group of Fermilab scientists in order to address the computing needs of fellow scientists in high energy physics and other concentrations. We have customized the operating system on the compute nodes to support diskless operation. This includes compilation of a custom Linux kernel, selection of the system's software packages, and modification of the file system structure to facilitate a read-only root file system.

Diskless compute node operation requires preparation of a custom Linux kernel. This involves selection of a kernel version with desired patches (source code modifications) and compilation of the kernel with custom options. Features added to the kernel include support for NFS root file system, swapping (virtual memory) over NFS, IP auto-configuration, and drivers for the network cards.

We obtain the source code for the kernel from the central repository for Linux kernel development, <http://kernel.org/>. The kernel version is chosen to correspond to the current kernel (version 2.4.21) used in the distribution of Scientific Linux 3.03 in order to ensure compatibility with other computing sites that use the kernel provided in the Linux distribution.

We apply a patch to the Linux kernel source code in order to add support for use of an NFS-mounted file system as virtual memory. This patch is provided by the sourceforge.net project NFS-Swapping4Linux [44]. In order to augment the RAM capacity on a computer, the system can temporarily move data from active RAM to a slower medium (usually a locally attached hard disk). This patch allows Linux to use a file on an NFS (remote) file system for this temporary staging of RAM data (virtual memory). There is a performance loss as compared to locally accessed virtual memory due to the latency of the network connection. However, this does not significantly affect our operation because our software rarely requires more RAM than is available on the compute nodes. The virtual memory mainly serves as a backup to prevent a system crash in the event that all RAM is used. The patch version must correspond to the kernel version. It is applied to the kernel source tree before kernel compilation using the Linux utility "patch."

⁹Scientific Linux is available online at <https://www.scientificlinux.org/>.

¹⁰The Red Hat Linux operating system is produced by Red Hat Inc. <http://www.redhat.com/>.

Before the kernel is compiled, some modules need to be added and compilation options changed from the default that the kernel is generally configured with. Many modules need to be directly compiled into the kernel (not as loadable modules) because the kernel will not have an accessible root file system through much of the boot process. We disable loadable modules for this reason.

The kernel option “NFS root file system support” allows us to mount a remote NFS-exported file system as the root file system (“/”). The option to enable virtual memory over NFS is automatically activated when the patch is applied. The compute node needs to get its IP address from the DHCP server before it can access anything over the network (including its root file system). The kernel option “kernel-level IP auto-configuration via DHCP” enables the kernel to obtain an IP address much earlier in the boot process than would normally be possible. Finally, in order to use the network, the kernel must include drivers for the network interface cards in the compute nodes. It is possible to include drivers for many different cards because the kernel will automatically detect the installed network card and use the appropriate driver.

In order to save memory we disable any kernel modules that the compute nodes do not need. These modules include support for disk storage, USB, sound cards, input devices, and video display devices.

Once properly configured, we compile the kernel into a compressed image file suitable for transmitting over the network during the compute node boot process.

3.3.2 The *BABAR* Software Framework

Software specifically written for and by *BABAR* collaborators is maintained in a central repository for easy access and use by the *BABAR* community. The software as a whole provides a framework to facilitate common computing needs such as data storage, access, analysis, and visualization as well as detector studies and simulations. The software is organized in a modular way that promotes code reuse and encapsulates functionality in discrete software units known as *packages*.

Packages

The package is the highest level modular unit of *BABAR* software. A single package contains a related set of functionality in the form of software classes.

Releases

A *release* is a defined set of specific packages and package versions that are designed and tested to work together. In general, each official *BABAR* software release contains all packages meant for general use. Releases are created for testing of new software, approved analysis use, Monte-Carlo simulation production, and other specific goals. Use of a release at a remote site (away from SLAC) requires the importation (copying) of all packages in the release followed by an installation procedure that checks for the existence and locations of all required libraries and support software not included with the release. This provides a way to ensure that the software environment is consistent between different *BABAR* sites.

3.3.3 Objectivity Database

There are a variety of levels of detail at which each data event in *BABAR* is accessible. The Objectivity [45] federated database originally served as the primary access facility for experimental data and MC events. The hierarchical storage system of the database allows access to event information at a variety of detail levels through a coordinating framework that provides for locking and data access functionality. Practical use of the system showed that most access to events occurred at a small range of detail levels, thus minimizing the advantages of the complicated system. The Objectivity database has since been phased out in favor of a more open and simple data access procedure using Kanga [46] data sets through the ROOT [47] framework. One exception is the use of Objectivity for the conditions and configuration databases in MC simulation production.

3.3.4 ROOT Framework

ROOT [47] is an object-oriented framework for scientific computing needs, and was developed at CERN as an improvement over the FORTRAN-based analysis systems in use. ROOT unifies data storage, manipulation, visualization, and characterization (fitting) into a single cross-platform solution. It has largely replaced the older Objectivity database for data storage and PAW [48] (the Physics Analysis Workstation) application for analysis within the *BABAR* collaboration. ROOT is based on ANSI C/C++ and features a C++ interpreter (CINT [49]) to enable quick development of scripts with access to the ROOT software interface.

RooFit

The *BABAR* collaboration has created a ROOT package, named RooFit [50], that standardizes and simplifies the production of fit-based analyses. RooFit provides a simple framework for unbinned maximum likelihood (ML) fitting and a library of PDF models. This package, along with ROOT version 4, is used for all ML fits in this analysis.

3.3.5 Other Software Packages

The Scientific Linux distribution includes a wide variety of software packages to support server or desktop use. The compute nodes require only a subset of software packages from the *Required Software for BABAR Computing* [51] list in order to run *BABAR* analysis and MC production software. We include the following software:

- CERNLIB [52] program library,
- GNU [53] utilities for Linux (gcc, gdb, gdiff, gfind, gmake, gtar, and patch),
- Perl [54] interpreted programming language,
- Tcl/Tk [55] shell script language.

3.3.6 File System Structure

We configure all compute nodes to access the same remote directory structure (via NFS) for their root file system. This requires the file system to be exported read-only to avoid conflicts with file creation and modification between compute nodes. However, some functions of the operating system require access to a writable file system for storage of temporary files and state information. Therefore, during the boot process, the compute nodes are configured to create a memory-resident file system for this purpose. The directories `/tmp/` (for temporary files) and `/var/` (for state information) are symbolically linked to this memory-resident file system. Since these file systems are located in RAM, all information is erased when the machine is re-booted. This is an advantage because it ensures that the compute nodes always boot into a known-good software configuration.

3.3.7 *BABAR* Production Software

Along with the primary MC production executable, there are a number of utilities provided by *BABAR* SP production.

The first step is to receive the SP request from the central database at SLAC. When a site has resources available, a script is executed that contacts the central database and requests jobs allocated to the site. If jobs are available, their details are retrieved and a directory is set up containing all the scripts and configuration files. Another script is then used to submit the jobs to the batch-queue system, with each job generating and reconstructing 2000 MC events.

When a job is finished, the resulting data file is verified to ensure that it is correctly formatted and that the collection contains the correct number of simulated events. It is then merged into a joint datafile along with other completed jobs from the same batch. The merged data collection is then uploaded to the central *BABAR* database at SLAC and is checked with a checksum to ensure that the transfer was error-free and contains the correct number of events. At this time, log files for the jobs are also uploaded to SLAC and stored for later debugging and quality-control purposes.

3.3.8 Custom Management Programs

We have also developed a number of utilities to help maintain efficient use of our computing resources and reduce the labor demands associated with running a large computing system. These utilities assist by increasing the level of monitoring and automation of the SP process and underlying computing infrastructure.

Monitoring

Online monitoring is an important tool for running an efficient computing cluster. Our monitoring is designed to maintain uptime and to resolve any problems within a few hours at most. We provide this with a combination of web and mobile phone accessible monitoring applications.

With the web monitor we track storage sub-system performance, network traffic, central server resource use, and the batch-queue as well as SP details such as number of queued SP jobs, status of data exports, and status of SP data validation checks. The monitor displays aggregate and individual performance statistics of the compute nodes in a color-coded format for quick evaluation (see Figure 3.4).

The second monitoring system is accessible through a mobile phone interface using the WAP [56] protocol. A process on the central server monitors a number of variables

BaBar Compute Cluster Status

updated Fri Jan 21 11:01:51 EST 2005

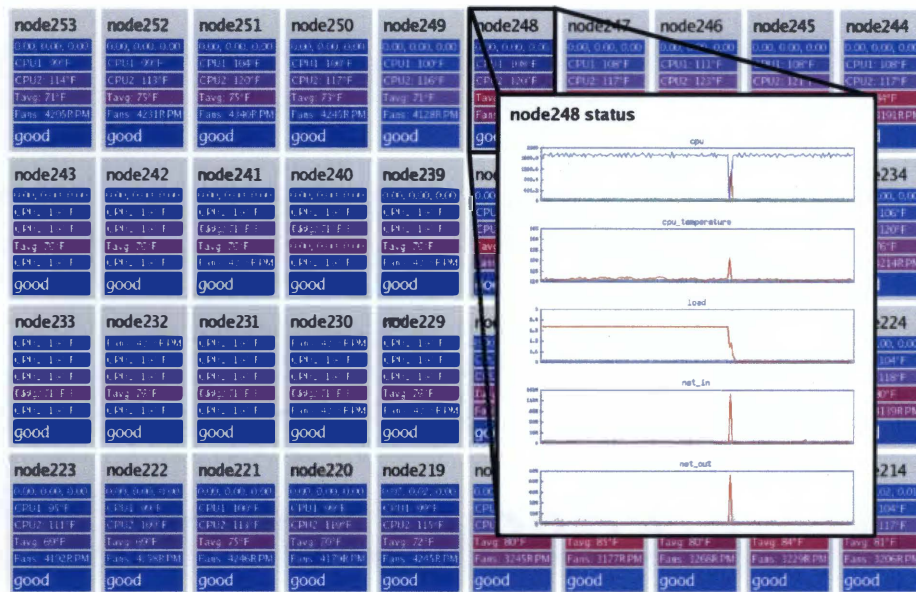


Figure 3.4: The online cluster monitoring software features a graph display of each compute node that shows the CPU load, internal temperatures, and CPU fan speed. Clicking on an individual node will show a report of CPU use, temperature, network activity, and virtual memory use over the last 24 hours. The status fields on the overall cluster view are color coded in order to quickly provide an indication of the activity of the cluster and to highlight problem areas.

such as hardware warnings, storage system use, and job failure rate. The system can send a mobile phone text message to the administrator in the event of problems that require immediate attention. The mobile phone interface includes reduced verbosity monitoring capability suitable for a small screen as well as limited service control (start and stop the batch queue). This system allows the administrator to quickly respond to error events that can later be addressed in more detail when a computer is accessible. All functions are protected by a user name and password combination with private key encryption.

Automation

The primary automation task for SP computing is the process of requesting and executing jobs, merging the resulting data, and exporting the data to the central SP database at SLAC. Using the Perl [54] programming language, we have developed an application (named “dwarf”) that provides this functionality as well as the ability to detect changes in the state of the computing system and respond appropriately. When the number of queued jobs reaches a certain threshold, dwarf requests more jobs from the site’s SP allocation, prepares the jobs, and submits them to the queue system. It monitors the progress of jobs and retries any failed job automatically. It also periodically runs the scripts that initiate the merge of data files and the export of merged data to the central database.

Additional SP-specific monitoring capability is built into dwarf to allow it to detect errors in specific machines (based on the rate of failed jobs) and to remove failed machines from the queue system. The problem with the machine can then be investigated by the SP administrator at a later time.

Documentation and source code for dwarf are available for use and modification [57].

3.4 Results

The computing cluster has proven to be efficient and reliable and has provided a significant service to the *BABAR* collaboration through the Simulation Production effort. The cluster came online in June 2003 with an initial 20 dual processor compute nodes. In January 2004, the capacity was further increased by 20 dual processor compute nodes with 20% higher clock speed CPUs. MC production has been continuous since coming online with the exception of periods during which no MC requests for *BABAR* were issued. As of January 2005, the cluster has produced over 250 million MC events: an average rate of nearly 1 million events per day. Figure 3.5 shows the time-integrated MC events produced by the cluster between June 2003 and January 2005.

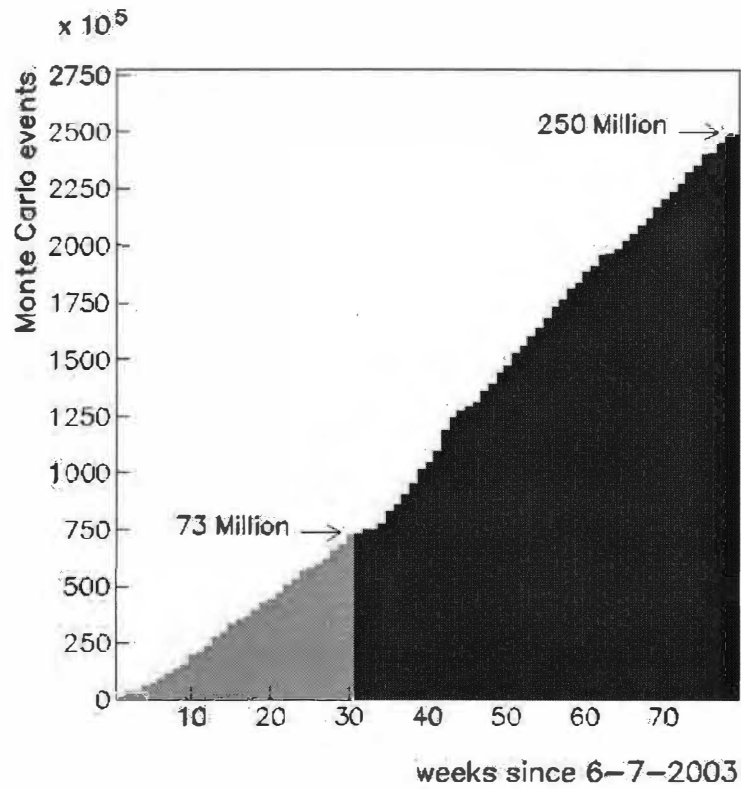


Figure 3.5: The Tennessee compute cluster has produced 250 million MC simulated events as of January 2005. This plots shows an upward kink in the MC production rate at about week 30 caused by the expansion of the cluster (which roughly doubled the available computing power). This increased production capacity was tempered somewhat by a software version upgrade (SP5 to SP6) that introduced a more computationally intensive simulation.

The cluster design makes efficient use of the available computing capacity. We quantify the CPU-use efficiency by measuring the percentage use of clock cycles on the compute node CPU during execution of the MC simulation. Latency in the network hardware and bottlenecks in the storage systems can result in the CPU waiting for resources (wasted clock cycles). We find the cluster makes more than 97% use of the compute node CPUs during MC simulations.

Downtime of the cluster has been very limited. Power outages have been very limited in the building with only one confirmed outage (for a few minutes) in two years. The UPS systems that protect each central server prevented their interruption, and the compute nodes automatically rebooted and continued operation once power was restored. Maintenance downtime has also been very limited due to the hot-swap hard disk controllers on the central file servers; a failed disk can be replaced without turning off the server.

In the time it has been operational, the cluster has provided a significant service to the *BABAR* MC production effort. MC production for *BABAR* is performed at over 15 sites around the world. Of these sites, our cluster ranks in the top 7 of individual universities by events produced since the cluster came online. Figure 3.6 shows the proportion of total MC events produced from all *BABAR* MC production sites.

3.5 Future Expansion

Future expansion of the cluster in the same facility is mainly limited by the available cooling capacity. Any expansion will require installation of a dedicated “chiller” system in the room (a closed-cycle refrigeration system with waste heat removed via the building’s chilled water system).

A second barrier to expansion is the bottleneck created by the current system of centralized file storage. There are many solutions to this problem in the form of cluster file systems. These systems distribute storage load and capacity among a number of separate machines while maintaining the centralized administration of a single storage server. We have conducted preliminary tests of the Parallel Virtual File System (PVFS) [58] and are currently investigating the Lustre¹¹ file system (a distributed cluster file system for Linux) as well as the SRM [59] and dCache [60] data management systems.

The network infrastructure is the last anticipated barrier to expansion. The increased network demand from hundreds of compute nodes will require new network switches capable of 1 Gbit/sec connections to all ports. Commodity network switches offer at

¹¹Developed by Cluster File Systems, Inc. <http://www.clusterfs.com/>

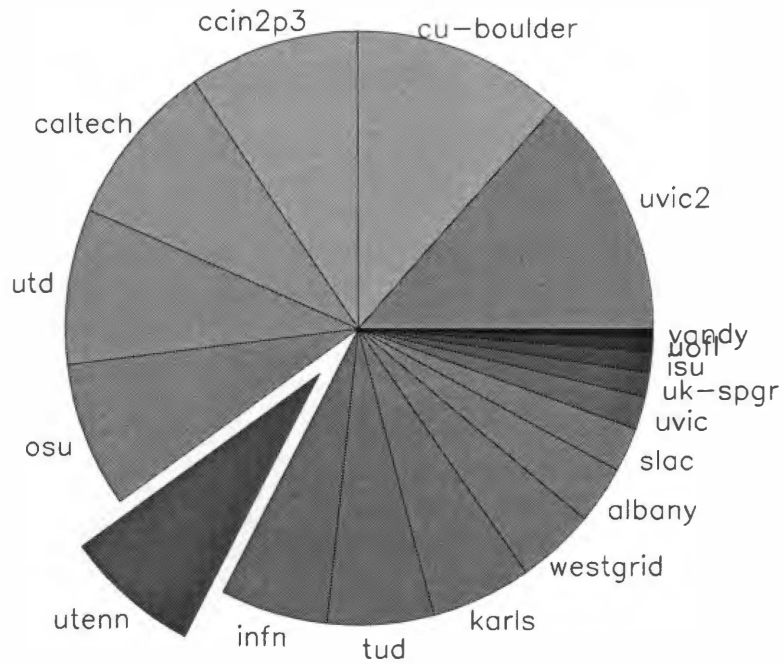


Figure 3.6: This shows the total proportion of MC events produced per *BABAR* SP site as of Jan. 2005. The contribution from the University of Tennessee *BABAR* group represents nearly 10% of total events produced and ranks our site among the top 7 university-based production sites.

most 48 ports, so the switches must have the capability of being “bridged” together via high speed optical connections. The external network connection must also be upgraded in order to accommodate the increased amount of data processed by the cluster. This upgrade will take the form of a 10 Gbit/sec fiber-optic connection directly to Fermi National Accelerator Laboratory (FNAL) (in anticipation of future collaborations) and in a local area network upgrade to 1 Gbit/sec.

Chapter 4

Particle Reconstruction

For a full reconstruction of a particle decay, the four-momenta (energy plus momentum) and origins of the decay products need to be estimated. In the following, we outline the reconstruction of tracks of charged particles starting with recorded raw charge signals in the SVT and DCH and the reconstruction of neutral particles with the EMC. We also detail the particle identification techniques used for separation of pions and kaons.

4.1 Charged Particle Reconstruction

The charged particle track reconstruction stage uses raw hits gathered from the SVT and DCH systems. These are produced mostly by electrons that are liberated by the primary charged particle in the material of the SVT and DCH and are collected in an electric field. A charge signal induced in one of several neighboring and well located anode strips of the SVT provide the average position of the particle trajectory within the silicon substrate on each side. With a signal on more than one strip, calculating the signal height weighted average improves the hit precision σ_x , along the line perpendicular to the strips, from $\sigma_x = \text{pitch}/\sqrt{12}$ (typical pitch is between $50\,\mu\text{m}$ and $100\,\mu\text{m}$). The induced charge signal on the well-located sense wire together with the reconstructed drift distance of the electrons constitutes a hit position along the particle's trajectory in the DCH.

During online processing, DCH pattern recognition and track finding algorithms [61] combine hits into candidate track segments. This initial information includes estimates of the distance of closest approach to the z axis (d_0), the track azimuthal angle (ϕ_0), and the starting time for the track. During offline reconstruction, other hits in the DCH consistent with these tracks are added. The tracks in the DCH are then extrapolated into the SVT volume, and hits in the SVT are added.

In the magnetic field within the detector, a charged particle's trajectory is a helix. Along this path, the particle can lose energy, undergo multiple scattering, and experience variations in the magnetic field strength. A proper description of the track requires five parameters for the helix. These parameters are the curvature (ω), the $x - y$ and z distances of closest approach to the z -axis (d_0 and z_0), the initial track azimuthal angle (ϕ_0), and the dip angle of the track from the transverse plane ($\tan \lambda$). A fit extracts the five helix parameters of each track using the least-squares technique (with modifications to account for material interaction and magnetic field distortion). For this reason, the fit requires a detailed model of the detector volume comprising the SVT and DCH. The fit begins with the track hits nearest the interaction point and fits the hits to a piecewise helix progressing in the forward (positive z) direction. The fit is then performed in the reverse direction in the same fashion. This piecewise method allows the fit to account for field variations, scattering, and energy loss in a simple way. The resulting values of the five helix parameters (for any given location along the track) are determined from a statistical combination of the fits in the forward and reverse directions. This procedure is a special *BABAR* implementation of the Kalman Fit algorithm [61].

The Kalman fit is then used to extract the five helix parameters from tracks found in both the SVT and DCH. Hits in the SVT that are not associated with known tracks are then fed into an SVT-only track fit. The final step attempts to combine tracks existing only in the SVT with those in only the DCH.

The positional resolution in the DCH cells provided from time-to-distance drift relations is typically between $100\,\mu\text{m}$ and $400\,\mu\text{m}$. Measurements of z and ϕ from the SVT have a typical resolution between $15\,\mu\text{m}$ and $40\,\mu\text{m}$.

Using the track description parameters, we can calculate the track momentum in terms of the particle's charge (q) and the z component of the magnetic field (B_z):

$$\vec{p}(l) = \frac{qcB_z}{\omega}(\cos(\phi_0 + \omega l)\hat{x} + \sin(\phi_0 + \omega l)\hat{y} + (\tan \lambda)\hat{z}). \quad (4.1)$$

The particle energy is calculated assuming different particle mass hypotheses.

4.2 Particle Identification (PID)

The measurement of dE/dx from the DCH dominates *BABAR* kaon identification for particle momenta below $700\text{MeV}/c$. PID for each track is achieved using an 80% truncated mean of at least 40 dE/dx measurements. This provides a 7% resolution on dE/dx . Figure 2.8 shows the momentum dependence of dE/dx in a varied particle sample.

For high momentum tracks, the DIRC's measurement of Cherenkov cone angle (θ_C) and Cherenkov photon number (N_γ) is *BABAR*'s primary PID information source. During reconstruction, photo-multiplier tube (PMT) hits are associated with tracks from the charged particle tracking system. When a sufficient number of Cherenkov photons is available, extraction of θ_C is possible. For each PMT photon hit, the possible emission angles and arrival times are calculated based on the space-time coordinates of the PMT signal. Timing and geometrical information are used to reduce both the photon background and ambiguity in photon-to-track association. Background is typically reduced by a factor of 40. A fit to the track associated photons extracts θ_C and the number of (signal and background combined) photons, N_γ .

Kaons that reach the Cherenkov detector have a minimum transverse momentum of 200 MeV/c; the Cherenkov threshold is $(0.92 \times 487) \text{ MeV}/c$. The resolution of θ_C scales like $1/\sqrt{N_\gamma}$ with a typical N_γ between 20 and 65 (depending on the distance the charged particle travels through the quartz bar of the DIRC). The average θ_C resolution is about 2.5 mrad and provides a pion to kaon separation of more than 2.2σ at 4.3 GeV (the highest possible momentum in 2-body B decays).

4.3 Photon Reconstruction

When a particle enters a crystal of the electromagnetic calorimeter (EMC), the resulting electromagnetic cascade causes a deposit of energy spread over many adjacent crystals (known as a *cluster*). More than one local energy maximum (a *bump*) within a cluster can arise due to the decay of a high energy neutral pion into two photons. The reconstruction algorithm searches for these characteristics by first identifying crystals with an energy deposit greater than 10 MeV. Using these as *seed* crystals it then adds crystals with $E > 1 \text{ MeV}$ that are adjacent to existing cluster member crystals with $E > 3 \text{ MeV}$. A crystal with a local maximum is defined as having $E'/E = (N - 2.5)/2$, where E' is the highest energy of the neighboring N crystals having $E > 2 \text{ MeV}$. The bump location is then determined with an iterative algorithm using the cluster crystal fractional energies. Finally, a cluster is associated with a particle track if its bump location matches a known particle trajectory. The cluster is otherwise assumed to belong to a neutral particle originating from the primary interaction point (IP) for the event.

A good cluster is required to have a total energy $E > 20 \text{ MeV}$. The lateral shape parameter λ_{LAT} [62] (a description of the lateral spread of the shower within the EMC crystals) must be less than 1.1. Finally, the cluster must have a lab frame polar angle θ_{lab} between 0.41 rad and 2.409 rad in order to be within the EMC volume.

4.3.1 π^0 Reconstruction

Each neutral pion is reconstructed from its two decay photons, registered as clusters in the EMC, with the initial assumption that the photons originate from the IP. A small percentage (less than 10%) of high energy pions result in EMC clusters that are too close to be differentiated as individual clusters. These *merged* pion photons are isolated as local maxima within the merged clusters, and the overlapping crystals are added with weights.

We require that the π^0 mass be within $30 \text{ MeV}/c^2$ of $m_{\pi^0} = 130 \text{ MeV}/c^2$ and that either the single π^0 energy be greater than 200 MeV or the merged π^0 energy be greater than 1 GeV . The π^0 invariant mass resolution is $6.9 \text{ MeV}/c^2$.

4.4 Event Reconstruction

4.4.1 ϕ Selection

We require that each event contains at least one $\phi \rightarrow K^+K^-$ signal. We accomplish this by first requiring each track be built from 12 or more hits within the DCH. From these tracks, at least one must be identified as a charged kaon by the PID system. We require that the track helix parameters satisfy $d_0 > 1.5 \text{ cm}$ and $|z_0| < 10 \text{ cm}$ for both candidate kaons. The track momentum is calculated from the helix parameters with the kaon mass hypothesis. We require the track momentum $|\vec{p}| > 10 \text{ MeV}/c$ and the transverse momentum (in the x-y plane) $p_T > 100 \text{ MeV}/c$. Finally, the combined mass of the two kaon candidates must be within $16 \text{ MeV}/c^2$ of the ϕ mass of $1.019 \text{ GeV}/c^2$ [23].

Kaon Selection

Kaon candidates are required to pass a likelihood-based selection process that uses dE/dx information from the SVT and DCH as well as Cherenkov angle information from the DIRC. The dE/dx information from the DCH is used for tracks with momenta less than $600 \text{ MeV}/c$, and SVT dE/dx information is also incorporated for tracks with momenta less than $500 \text{ MeV}/c$. The dE/dx versus momentum plot (Figure 2.8) shows the separation of particles into bands by their mass values.

Above $600 \text{ MeV}/c$, the dE/dx separation degrades due to relativistic energy loss (causing the separate bands to merge). Above $2.5 \text{ GeV}/c$, the bands again reach a separation of about 2σ in dE/dx . However, above $600 \text{ MeV}/c$ the particles have sufficient momentum to produce Cherenkov photons in the DIRC. In this momentum range, the DIRC becomes the important factor for a likelihood selection. The Cherenkov angle is compared

to the particle momentum for the three mass hypotheses (Figure 2.7), and the likelihood for each measured particle and hypothesis is calculated based on the theoretical expectation:

$$\cos(\theta_C(p)) = \frac{cm}{np} . \quad (4.2)$$

The information from the three PID detectors is combined in terms of a likelihood ($L^i = l_{\text{SVT}}^i l_{\text{DCH}}^i l_{\text{DIRC}}^i$), where l^i is one of the five possible particle types (electron, muon, pion, kaon, or proton). This allows a continuous application of all information. The dE/dx likelihoods l^i (for the SVT and DCH) are Gaussian probability functions with respect to the expected (calibrated mean) $(dE/dx)^i$ value at that particle momentum: $\text{Gaussian}(dE/dx, (dE/dx)^i, \sigma(dE/dx))$. The DIRC likelihood is the product of a Gaussian probability (for the expected Cherenkov angle θ^i) and a Poissonian probability (for the observed number of Cherenkov photons):

$$l_{\text{DIRC}}^i = \text{Gaussian}(\theta, \theta^i, \sigma(\theta)) \cdot \text{Poissonian}(N_\gamma, N_{\text{expected}}) , \quad (4.3)$$

where the poissonian mean is the expected number of photons N_{expected} (derived from a table dependent on track angle of incidence, momentum, direction (charge), and DIRC bar number). The *a priori* probability for each of the particle hypotheses is 0.2.

The requirements for a certain particle type selection are based on likelihood ratios. For the kaon selection it is $\kappa = L_K / (L_K + L_\pi + L_{\text{proton}})$ (since only the hadrons are relevant). Based on a momentum dependent cutoff for κ , a certain kaon selection efficiency and particle mis-identification rate (most from pion mis-identification) can be achieved. The particular choice for the kaon selection in this analysis is displayed in Figure 4.1. We require that only one charged kaon from the ϕ decay in the $B^0 \rightarrow \phi K_S^0$ channel be identified. The average efficiency for the requirement is 90%, while the average mis-id rate is below 2%.

For a typical charged kaon from the ϕ in the decay $B \rightarrow \phi K$, the selection has an average efficiency of about 85% and a pion mis-identification rate of less than 2%. This is determined from the control samples of $D^* \rightarrow D^0 \pi$, $D^0 \rightarrow K \pi$ events (where the kaons and pions are identified via the decay kinematics). The detailed performance of the kaon selection has been characterized as a function of the laboratory momentum and can be seen in Figure 4.1.

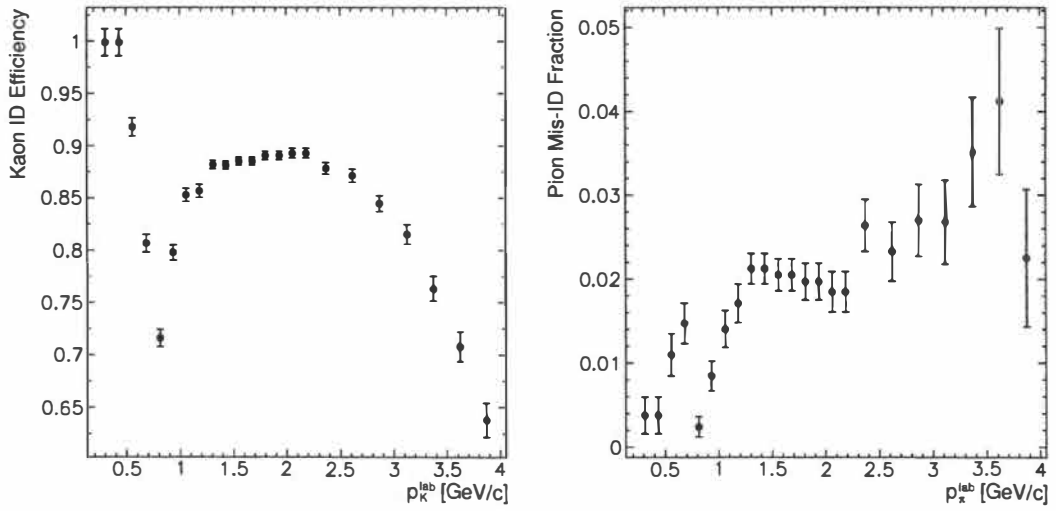


Figure 4.1: This shows the identification efficiency (left) of the charged kaon selection as a function of the kaon laboratory momentum p_K^{lab} and the fraction of charged pions misidentified (mis-ID) as kaons as a function of the pion laboratory momentum p_π^{lab} (right). The error bars represent statistical uncertainties in the control sample of kaons and pions from $D^* \rightarrow D^0\pi$, $D^0 \rightarrow K\pi$ decays. It is not required that the particle track traverses all particle identification detectors; notice that the DIRC solid angle coverage is only 87%.

4.4.2 K_S^0 Selection

We reconstruct the decay $K_S^0 \rightarrow \pi^0 \pi^0$ from its final state radiation. The four decay photons of the neutral pions are detected as either separate or merged bumps in the EMC. The K_S^0 decay point is initially chosen at the origin (IP). This causes an overestimation of the angles between the photons and leads to an invariant mass that is too low. The K_S^0 decay vertex is then determined by moving it along the line of the initial momentum, using the decay position of the ϕ as a geometrical constraint, while minimizing the difference with the expected invariant mass of the K_S^0 [5]. This fit uses the *WalkFit* [63] vertexing algorithm.

We require the combined mass of the two π^0 to be within the range $340 \text{ MeV}/c^2 < m_{\pi^0 \pi^0} < 610 \text{ MeV}/c^2$ and the combined energy within the range $0.8 \text{ GeV} < E_{K_S^0} < 6 \text{ GeV}$. The flight distance $c\tau$ is defined as the distance between the vertices of the ϕ and K_S^0 in three-dimensional space. We require $c\tau > 0.1 \text{ mm}$. The back-to-back decay of the ϕ and K_S^0 in the center of mass (CM) frame of the B candidate places a strong constraint on the momentum direction of the K_S^0 . We define an angle between the line joining the ϕ and K_S^0 vertices, in a plane perpendicular to the beam axis (α_{xy}), that represents the flight direction and momentum direction of the K_S^0 in the laboratory frame. We require $\cos \alpha_{xy} \geq 0.999$.

Chapter 5

Backgrounds

5.1 Continuum Background Suppression

The major background source of $B^0 \rightarrow \phi K^0$ candidates is ϕ production by $e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ processes. These are known as continuum events.

An effective way to suppress continuum background is to make use of variables that are sensitive to the jet-like topology (shape) of these background events. In a signal event, e^+e^- produce $B\bar{B}$ via the $\Upsilon(4S)$ resonance. In the $\Upsilon(4S)$ rest frame, the B has low momentum and the decay is fairly isotropic (Figure 5.1(a)). There is also a low correlation between the directions of decay products from each B . In a light quark continuum event, the event shape has a two-jet structure (Figure 5.1(b)); there is a strongly preferred direction characterizing the whole event [64]. In such events, the hadrons which form a B candidate are usually found in one of two nearly back-to-back jets. This background is suppressed by placing requirements on variables that are sensitive to this event topology.

5.2 B Meson Decay Backgrounds

For estimates of backgrounds from $e^+e^- \rightarrow B^0\bar{B}^0$ and $e^+e^- \rightarrow B^+B^-$ decays, we use samples of generic Monte-Carlo (MC) data. All reconstruction procedures and pre-selection criteria are applied and the resulting events are categorized according to decay channel. If any significant background source is identified, it is then isolated for study using signal MC for the particular channel in question. Several times more MC, corresponding to the luminosity in data, must be generated in order to estimate the contribution of fake B signals.

Table 5.1 lists B background contributions from 144 million $B^0\bar{B}^0$ MC and 88 million B^+B^- MC events as well as 578,000 $B^0 \rightarrow \phi K_S^0, K_S^0 \rightarrow \pi^0\pi^0$ signal MC events. The

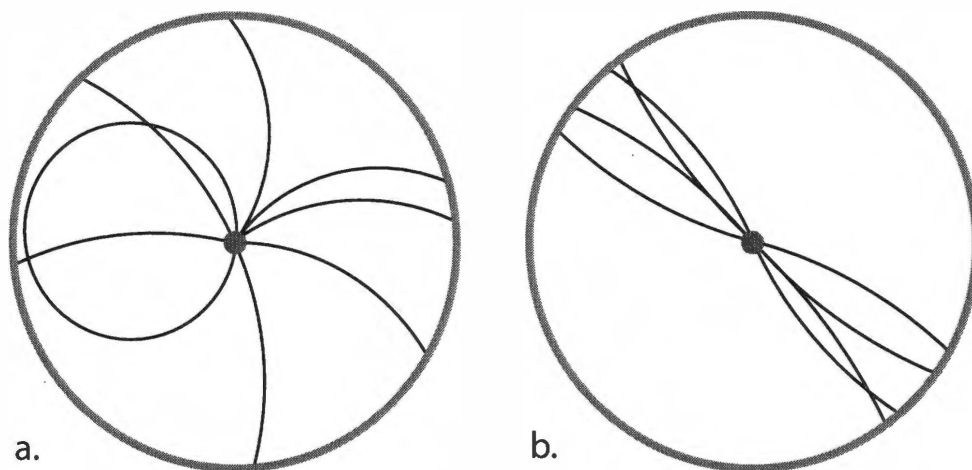


Figure 5.1: This diagram shows the general distribution of decay tracks from $B\bar{B}$ events (a) and continuum background events (b) in the (r, ϕ) CM frame. Continuum background event particles travel in *jets* which group around opposing axes (the thrust axes) while $B\bar{B}$ events are isotropically distributed in ϕ .

Table 5.1: The B background estimate from Monte-Carlo data sets shows a very small background contribution from generic B decays. Estimates of the signal yield in data are made based on either a sample of signal MC (first row) or the sample of $B^0\bar{B}^0$ generic MC (row four). The difference in these two values is attributed to the much higher uncertainty in the second value due to smaller sample size.

Event Mode	Events generated	Events passed	Yield in 200 fb
signal MC	578k	83755	18
SXF signal		15078	3
$B^0\bar{B}^0$ MC	144M	80	
signal		72	21
background		8	2
$f^0(980)K_s^0$		3	1
ϕK^*		1	0
$\phi\eta$		2	0
$\phi\pi^0$		1	0
$D^-\eta\rho^+\pi^+\pi^-$		1	0
B^+B^- MC	88M	0	0

resulting data sets contained 72 signal events, eight $B^0\bar{B}^0$ background events, and zero $B^+ B^-$ background events. We conclude that there is no significant background contribution from either $B^0\bar{B}^0$ or B^+B^- decays. Thus, these channels will not require a more detailed treatment to be adequately accounted for.

Chapter 6

Event Selection

6.1 Event Variables

The final event characterization is based on a set of discriminating variables that contain the kinematic and topological information necessary to separate signal from background events. We use four event variables: the beam-energy substituted mass (m_{ES}), energy difference (ΔE), Helicity ($|\cos \theta_H|$), and Fisher Discriminant ($\mathcal{F}$).

6.1.1 m_{ES}

The beam-energy-substituted mass is defined as

$$m_{ES} = \sqrt{(s/2 + \vec{p}_T \cdot \vec{p}_B)^2 / E_T^2 - \vec{p}_B^2}, \quad (6.1)$$

where q_T is the four-momentum of the $\Upsilon(4S)$, $s = q_T^2$ is the square of the center-of-mass (CM) energy, $\vec{p}_T$ and $\vec{p}_B$ are the three-momenta of the $\Upsilon(4S)$ and the B in the laboratory frame, and $E_T = q_T^0$ is the energy of the $\Upsilon(4S)$ in the laboratory frame.

6.1.2 ΔE

The beam-constrained energy of the candidate B meson is obtained as follows:

$$E_{bc} = \frac{s - 2\vec{p}_i \cdot \vec{p}_B}{2E_i}. \quad (6.2)$$

This is then used to calculate the energy difference

$$\Delta E = E_B - E_{bc}. \quad (6.3)$$

$(E_i, \vec{p}_i)$ is the four momentum of the initial state, and $(E_B, \vec{p}_B)$ the four momentum of the candidate B meson, both calculated in the laboratory; E_{bc} results from the assumption that we have particle-antiparticle production with $\vec{p}_B$ being the net three-momentum of the decay products of particle and antiparticle.

6.1.3 Helicity

Angular momentum conservation in the decay of a pseudo-scalar ($J = 0$) B meson to the vector ($J = 1$) ϕ and pseudo-scalar K requires that the ϕ be produced with Helicity zero. The subsequent decay of the ϕ into two kaons then yields a distribution of the cosine of the decay angle of the kaon, in the rest frame of the ϕ , with respect to the direction of the ϕ in the B frame, $\cos \theta_H$, which shows a characteristic $\cos^2 \theta_H$ behavior. Continuum background is flat in this distribution.

6.1.4 Fisher Discriminant

The primary tools for separation of signal from continuum background events are the topological (event shape) variables Thrust angle, B production angle, and Legendre moments ($\mathcal{L}_0$ and $\mathcal{L}_2$).

Thrust Angle

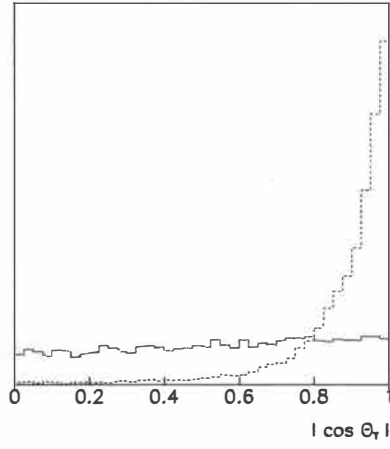
The continuum events form two back-to-back jets. The Thrust axis is defined as the direction in the $\Upsilon(4S)$ rest frame along which the sum of the longitudinal momentum components of particles in the jets is maximized. We consider the cosine of the angular difference between the thrust axes of the B candidate and the other charged and neutral particles of the event, $|\cos \theta_T|$, which peaks at one for continuum events. For B events, the two mesons are formed nearly at rest and the direction of their decay products are uncorrelated. This results in a spherical shape of events and an uniform distribution of $\cos \theta_T$ (Figure 6.1(a)).

B Production Angle

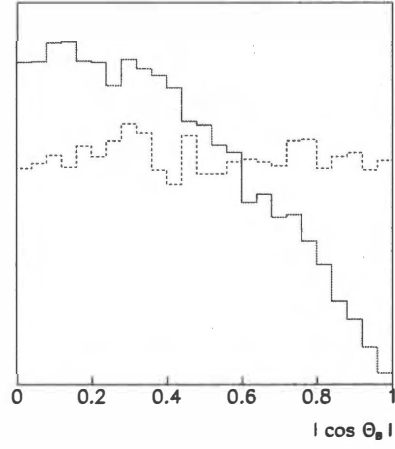
The B production angle (θ_B) is formed by the B candidate with respect to the beam axis (z-axis) in the $\Upsilon(4S)$ rest frame. This angle has a $\cos \theta_B$ distribution for B meson decays. In contrast, for continuum events, this angle is expected to have a nearly flat distribution (Figure 6.1(b)).

Legendre Moments

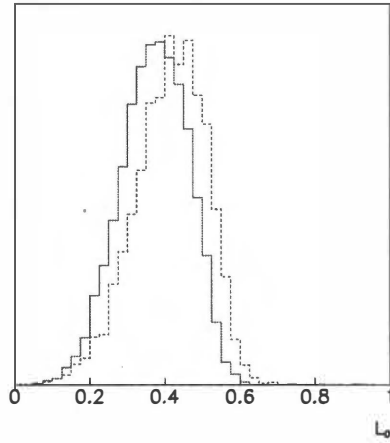
The sphericity of events can be further exploited by using the energy flow in 9 consecutive angular cones along the B candidate thrust axis. The shape of the flow in the 9 bins is characterized with a momentum-weighted Legendre polynomial: $L = \sum_j p_j \mathcal{L}_j$, where $\mathcal{L}_j$ is the j th Legendre polynomial [65]. The Legendre moments with highest discriminating power, $\mathcal{L}_0$ and $\mathcal{L}_2$, are used as topological variables (Figures 6.1(c) and 6.1(d)).



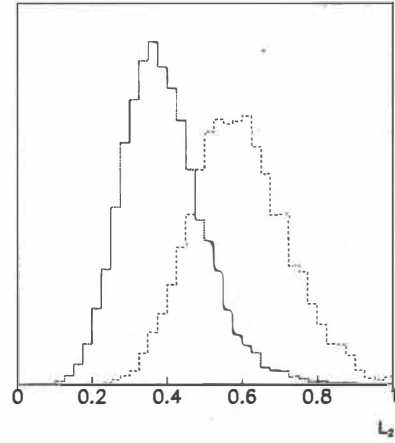
(a) Thrust angle



(b) B production angle



(c) $\mathcal{L}_0$



(d) $\mathcal{L}_2$

Figure 6.1: Distributions of the four topological variables used in continuum background suppression for signal (solid red) and background (dashed blue) including the Thrust angle (a), B production angle (b), $\mathcal{L}_0$ (c), and $\mathcal{L}_2$ (d) with arbitrarily normalization.

Table 6.1: Correlation coefficients for each combination of event shape variables used for continuum background suppression

	$\mathcal{L}_2$	θ_T	θ_B
signal			
$\mathcal{L}_0$	0.0428	-0.0027	-0.0104
$\mathcal{L}_2$		0.7675	0.0064
θ_T			0.0064
background			
$\mathcal{L}_0$	0.5742	0.1978	-0.0248
$\mathcal{L}_2$		0.5784	0.0030
θ_T			-0.0094

Combined Usage

Each of these variables provides significant background suppression. For example, the requirement $|\cos \theta_T| < 0.9$ suppresses continuum background by a factor of 3 while the signal is reduced by only 10% (Figure 6.1(a)). However, these variables are strongly correlated (Table 6.1). Therefore, requirements for the individual variables will result in a stronger signal reduction as compared to an optimized cut in the multi-variable space. This also makes the independent use of the variables in a likelihood fit impractical.

For each event, we define a new variable $\mathcal{F}$, which consists of a linear combination of the four event shape variables (x_{1-4}) with weights (A_{1-4}):

$$\mathcal{F} = \sum_{i=1}^4 A_i x_i. \quad (6.4)$$

This new variable is known as a Fisher Discriminant ($\mathcal{F}$) [66]. Once optimized to separate continuum events from B decays, it allows one to apply a quasi multi-variable criteria for the purposes of continuum suppression.

The separating power is optimized by first setting the Fisher coefficients A_i to random values. Then two normal distributions are created by applying the Fisher calculation to data sets consisting of either signal MC or sideband data (data excluding ΔE and m_{ES} values close to the expectation for B -decays that are representative of background). The Fisher coefficients are then varied in order to produce distributions for signal and background that maximize the distance between their means. These optimized coefficients are of universal use for continuum background suppression.

In chapter 7 we explore the use of multilayer perceptron (MLP) algorithms, which present potential for increased separating power as compared to the Fisher Discriminant.

6.2 Final Event Selection

For the event pre-selection, we require

- $m_{\text{ES}} > 5.21 \text{ GeV}$,
- $|\Delta E| < 0.15 \text{ GeV}$,
- $|\mathcal{F}| < 4.0$,
- $|\cos(\theta_T)| < 0.9$.

This reduces continuum background substantially while ensuring that the signal remains nearly 100% contained within the remaining events.

6.2.1 Multiple Candidates

After the pre-selection, we find multiple reconstructed $B^0 \rightarrow \phi K_S^0$ candidates in a single event. About 50% of all events contain two candidate signal events that satisfy our criteria. We expect that this rate decreases with more stringent requirements. We make a choice based on closeness to the decay particle hypothesis in terms of a χ^2 probability:

$$\chi^2 = \left(\frac{m_{K_S^0} - m_{\pi^0\pi^0}}{\sigma_{m_{\pi^0\pi^0}}} \right) + \left(\frac{m_\phi - m_{K^+K^-}}{\sigma_{m_{K^+K^-}}} \right) \rightarrow P(\chi^2). \quad (6.5)$$

We require a minimum probability of 0.005. It turns out that the K_S^0 probability alone resolves 85% of the candidates correctly, while the ϕ probability alone is 58%.

6.2.2 Photon Association Study

The π^0 daughter particles of the K_S^0 each decay into a pair of photons that register in the EMC. We use these photon hits to reconstruct the K_S^0 vertex and energy. In this decay channel, there is a significant number of signal events that exhibit an incorrect reconstruction of the π^0 daughter particles of the K_S^0 . These reconstruction errors (known as self cross-feed, or SXF, events) can occur three ways:

1. The event contains four reconstructed photons from the K_S^0 that are descendant from the pair of neutral pions, but these photons are not correctly assigned to the pion parents.
2. The event has fewer than four photons correctly associated with the neutral pions.

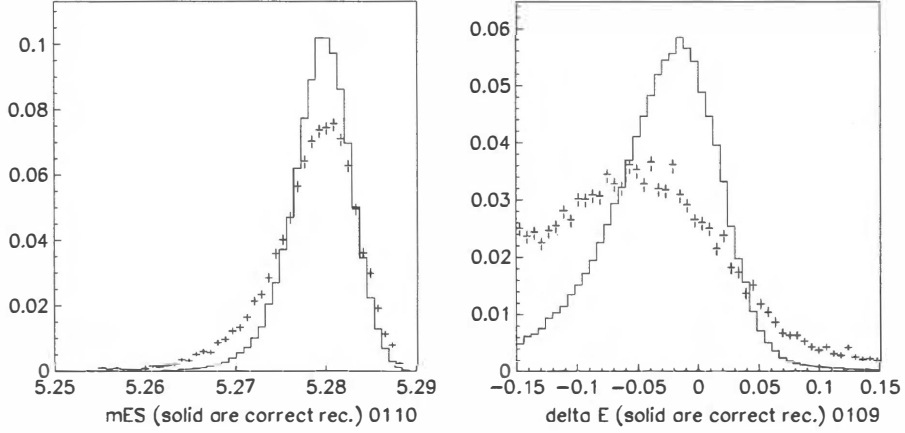


Figure 6.2: Signal Monte-Carlo shows significant differences in m_{ES} (left) and ΔE (right) between well-reconstructed signal events (solid) and SXF signal (error bars). Both histograms for SXF and good signal are normalized to the same number of events.

3. The event contains other reconstruction errors that preclude correctly reconstructing the π^0 photons.

Incorrect photon association distorts the reconstructed decay vertices and therefore the photon direction. This changes the reconstructed energy, which depends on the photon polar and azimuthal angles. The distributions for SXF events in the m_{ES} and ΔE event variables will be distorted (Figure 6.2).

We isolate and parameterize events with incorrect photon association using signal Monte-Carlo. We compare the angular components of the reconstructed photon momenta with the originally generated momentum components. We require that the residual squared $\chi^2 < 0.02$, where $\chi^2 = (\theta_r - \theta_g)^2 + (\phi_r - \phi_g)^2$. We then define SXF events to be those that do not satisfy the χ^2 criterion or have photons that are incorrectly grouped to pion parents. With this definition, 15.3% of signal events are classified as SXF events. The SXF categories are listed in Table 6.2.

6.3 Results

In Table 6.3, we present the efficiencies for the different criteria of event selection. From 578,000 signal MC events, 28.8% of events pass the reconstruction and pre-selection. The remaining efficiencies are calculated with respect to the number of events that pass the pre-selection. Due to correlations among the parameters, the product of all partial

Table 6.2: We used 578,000 signal MC events to determine the fraction of mis-association of π^0 photons. We find that 15.3% of events have photons that are incorrectly associated with the π^0 parent, are missing one or more photons, or have other reconstruction errors that preclude correctly assigning the π^0 photons.

Photon assoc. type	Number of events	Rate
correct assoc.	83755	84.7%
incorrect assoc.	210	0.210%
photons = 3	1962	1.99%
photons = 2	291	0.29%
photons = 1	60	0.060%
photons = 0	11408	11.5%
other	1147	1.16%

efficiencies is lower than the result of the combined application of all criteria. After all criteria, the total reconstruction efficiency is 17.1% with a negligible uncertainty, due to MC statistics, of 0.3%.

Table 6.3: This analysis of pre-selection and event selection requirement efficiencies uses signal Monte-Carlo. All efficiencies following pre-selection are with respect to the number of events passing pre-selection.

Item	Events	Eff.
generated (SP6)	578000	
pre-selection	166643	28.8%
m_ϕ	158707	95%
m_{ES}	166598	100%
ΔE	149759	90%
thrust	148359	89%
m_{K_s}	166481	100%
Δt	163715	98%
$\sigma_{\Delta t}$	158016	95%
K_s^0 probability	146614	88%
K_s^0 lifetime	159138	95%
$\cos \alpha_{xy}$	157736	95%
total efficiency		17.1%

Chapter 7

Use of Multi-layer Perceptron for Background Suppression

In the following, we study an alternative approach to continuum background suppression before resuming the branching fraction measurement.

A *perceptron* is a classification, or decision-making, algorithm that learns concepts by repeatedly studying examples presented to it. It is a type of artificial neural network invented in 1957 at the Cornell aeronautical laboratory by Frank Rosenblatt [67]. It was quickly proven that linear/single-layer perceptrons could not be trained to recognize many classes of nonlinear-separable patterns [68]. Only by 1980 was it discovered that multilayer perceptron (MLP) algorithms overcome such limitations.

Through the use of example data (consisting of input and desired output), the MLP can be optimized (trained) to return an approximation to the desired result on any future (unknown) input data. There is evidence that analyses of B decays improve by replacing the current Fisher Discriminant continuum background suppression with the more complex MLP [69]. In this investigation, we determine the suitability of MLP background suppression for our specific analysis and compare the results to other established methods. We use the PAW [48] MLP implementation, MLPFIT [1].

7.1 The Multi-layer Perceptron

For our purpose, the MLP combines two or more discriminating variables into a single reduced set of variables. Ideally, this output variable still contains all the desired information from the constituent inputs and their correlations (if any).

The MLP is based on solid mathematical grounds [70]. Figure 7.1 shows an MLP with three input variables and one output variable. Input quantities are processed through

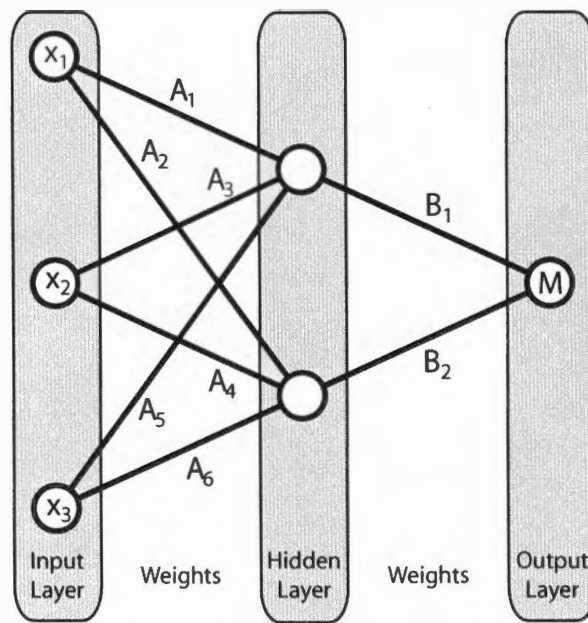


Figure 7.1: This diagram of a Multi-layer Perceptron (MLP) consists of an input, an output, and one hidden layer. Each layer contains one or more nodes. Each node is connected to every other node in adjacent layers by lines to represent the weights used in the MLP calculation. The example MLP network consists of three input nodes, two hidden layer nodes, and one output node. Equation 7.4 shows the equivalent mathematical notation for this specific MLP network.

successive layers of “neurons” or “nodes” (represented by circles in the diagram). The nodes on the left of the diagram represent the inputs and the single node on the right represents the output of the MLP. Vertical columns of nodes in the diagram are called *layers*. Each node in a given layer is connected to all other nodes in adjacent layers by lines. The connecting lines represent the path that a value takes as it propagates through the perceptron network.

Most perceptrons contain at least one extra layer of nodes in addition to the input and output layers. These additional layers are referred to as *hidden* layers. With no hidden layer, the MLP results in a simple linear combination of input values (similar to the the Fisher discriminant). All problems which can be solved with an MLP can be solved with only one hidden layer. However, the addition of a second hidden layer can reduce the time required to train the network.

In Figure 7.1, three input values start at the left side of the network. For each node (j) in the next hidden or output layers we calculate a linear combination (u_j) of all node values (x_i) from the previous layer:

$$u_j = A_{0j} + \sum A_{ij}x_i . \quad (7.1)$$

The lines in the network represent the weights A_{ij} and are determined during the training cycle from examples by minimizing the deviation, or *Error*, from a desired output. Given a $1 \times n$ matrix of desired output values ($\mathbf{D}_i$) and a $1 \times n$ matrix of the current output values from the MLP ($\mathbf{X}_i$), the deviation is defined as the l^2 -norm of the difference matrix $\mathbf{A} = \mathbf{D} - \mathbf{X}$:

$$\text{Error} = |\mathbf{A}| = \sqrt{\sum_{i=1}^n \mathbf{A}_i^2} . \quad (7.2)$$

Each output from a hidden layer (y) is modified according to a *transfer* function $y' = f(y)$. The resulting values are then used as input for the nodes in the next layer of the network. This continues until the final output layer of nodes is reached and output values are provided.

For this analysis we use a sigmoid as the transfer function:

$$f(y) = \frac{1}{1 + e^{-y}} . \quad (7.3)$$

An MLP with one hidden layer performs a linear combination of sigmoid functions of the inputs. The use of the sigmoid function is motivated by two theorems:

1. A linear function of sigmoids can approximate a continuous function of one of more variables. This allows one to obtain a continuous function to describe a finite set of points when no underlying model is available.
2. Trained with desired answers of 1.0 for signal and 0.0 for background, the approximated function is the probability of signal.

From the distribution of the variables for signal and background examples, we can extract a probability value optimized for signal efficiency and background suppression. This variable can also be introduced in the maximum likelihood fit.

For the example MLP (Figure 7.1) with three inputs, one hidden layer with two nodes, and one output, the explicit mathematical form of the network output is

$$M = A_{00} + f(A_{01} + A_1x_1 + A_3x_2 + A_5x_3)B_1 \\ + f(A_{02} + A_2x_1 + A_4x_2 + A_6x_3)B_2$$

or

$$M = A_{00} + \sum_{i=1}^2 B_i f(A_{0i} + \sum_{k=1}^3 A_{ik}x_k). \quad (7.4)$$

It is possible to write out the explicit mathematical form of any given MLP, although the results become unwieldy very quickly for all but the most simple networks. However, this explicit form is useful when comparing the MLP to the Fisher Discriminant.

7.2 Comparison with the Fisher Discriminant

The Fisher Discriminant can be described as a linear form:

$$F = \sum_{i=1}^n A_i x_i, \quad (7.5)$$

where n is the number of input variables for the discriminant, and the A_x coefficients are the input-specific weights applied to each input. An MLP network consisting of n inputs, no hidden layers, and a single output simplifies to

$$M = A_0 + \sum_{i=1}^n A_i x_i. \quad (7.6)$$

This is identical to the Fisher Discriminant form with the arbitrary normalization constant A_0 set to zero. Thus, the Fisher and MLP discrimination results can be compared within the same software framework.

7.3 Network Topology

Unlike the Fisher Discriminant, the MLP requires the specification of one or more hidden layers (each of which contains two or more nodes). Hidden layers serve to quantify correlations between various combinations of the input variables. Where such correlations exist, the hidden layers add discrimination power as compared to the Fisher Discriminant. However, when no correlations exist there is no benefit from the added complexity associated with the MLP. For this reason, it is important to experimentally determine the optimum network complexity for each specific MLP instance. In our analysis, we test various network sizes (number of nodes per hidden layer) using both one and two hidden layers. Once an optimum network (having the lowest l^2 -norm) is determined, we use the network in our comparison with the Fisher-equivalent MLP network.

7.4 Network Training

An important step in MLP use is the determination of the network weights A_{ij} . The weights are generally determined algorithmically in a process known as *training*. Training requires a data set containing a large number of input variable values and the output values desired from the MLP. The training data set should be large enough to ensure reasonable coverage of the expected variable phase space. For background separation, we choose a data set consisting of signal events (from MC), with a desired output value of 1.0, and an equal number of background events (from off-resonance data), with a desired output value of 0.0.

Along with the training data set is an identically prepared data set (containing unique data) which is used to gauge the progress of the training. The procedure consists of training the MLP iteratively using the training data and then testing the resulting MLP using the testing data. The testing is performed on a separate set of data in order to combat the tendency of the MLP to become *over-trained*: the specific training sample “signatures” can become encoded in the weights of the MLP when the complexity of the MLP (the number of weights) gets large compared to the number of unique training samples. Although an over-trained network will perform well on the data it was trained against, it may not perform well on other data.

Training is considered to be finished when the l^2 -norm of the error matrix (Equation 7.2) is minimized. Throughout this analysis, separate training and testing data sets are used, although relatively small MLP sizes and large number of training samples make over-training unlikely.

7.4.1 Training Methods

A number of training algorithms exist, all of which iteratively seek a minimization of the network Error (Equation 7.2). For each training iteration (known as a training *epoch*), the Error is determined for the current state of the MLP by evaluating the MLP output for the testing data. Based on a plot of the network Error versus epoch one can determine when the network is optimally trained, namely, when the testing sample Error reaches a minimum.

The most common training algorithm is known as stochastic minimization, an application of Robbins-Monro stochastic approximation [71] to MLP optimization. This approach proceeds by slightly modifying all MLP weights after each training example based on the first derivative of the network Error versus weight value. This method is not the most CPU-time efficient, but we find that it produces consistent results with low final network Error as compared with the other training algorithms available. This method is preferred for this analysis.

7.4.2 Training Procedure

We use equal amounts of off-resonance background data and Monte-Carlo simulated signal data for the training and testing of the MLP networks. One might expect that the background to signal ratio for testing would reflect the ratio expected in the data analysis. However, since the ratio of background to signal events is expected to be very large, the MLP networks would effectively only be trained on background data. Therefore, by training on equal parts background and signal, the network gets equal exposure to both and is thus not biased due to the training. We verify this by testing MLP networks on control data sets with high background to signal ratios.

The inputs for the MLP are identical to that used for the Fisher Discriminant. Each variable is first mapped onto the range $[0, 1]$. It is not critical exactly how this is accomplished, however a monotonic map is best since it avoids possible loss of information.

It is possible that one or more of the MLP training algorithms is dependent on the ordering of the data. We avoid this by passing the background and signal data sets separately through a Fisher-Yates shuffling algorithm [72] to randomize the order before splitting them into equal sized sets for training and testing.

We train the network for a fixed number of epochs which we have previously determined will result in a sufficiently minimized network Error. Using a real-time plot of the training and testing data set Errors versus training epoch, we can determine when the MLP is optimally trained. The possibility of an over-training condition is indicated

when the testing sample Error begins to increase after reaching a minimum. Figure 7.2 shows the progression of MLP error versus training epoch for a number of network sizes.

Once the MLP network is optimally trained, the weights obtained from the training are used to calculate the network output according to Equation 7.4. We produce a histogram of the network output from the testing sample events in order to visually verify that the network is functioning properly. Figure 7.3 shows the network output from an MLP having five inputs and a single hidden layer with five nodes.

7.5 Results

We investigated the suitability of MLP background suppression techniques for the decays $B^0 \rightarrow \phi K_S^0$, $K_S^0 \rightarrow \pi^0 \pi^0$ (K_S^0) and $B^0 \rightarrow \phi K_L^0$ (K_L^0).

7.5.1 $B^0 \rightarrow \phi K_S^0$, $K_S^0 \rightarrow \pi^0 \pi^0$ Dataset

For the $B^0 \rightarrow \phi K_S^0$, $K_S^0 \rightarrow \pi^0 \pi^0$ data sample, we prepared 4000 events of Monte-Carlo simulated signal events and 4000 off-resonance events taken at 40 MeV below the $\Upsilon(4S)$ resonance energy. We split each data set into 1000 events for training and 3000 events for testing. In order to determine the optimum network size for this application, we investigated networks with between one and nine hidden layer nodes. We used the stochastic minimization algorithm to train each network for 10,000 epochs. Each network was trained ten times, each time recording the minimum network Error achieved. We combined these ten minimum Errors to calculate the absolute minimum Error, the average minimum Error, and the standard deviation of the minimum Error for each network size (Table 7.1).

The network with one hidden layer node is equivalent to the Fisher discriminant. The MLP with five hidden layer nodes has the highest background separation performance. However, it represents less than a 1% performance improvement over the Fisher. We therefore decided that this does not justify the computational overhead and used the Fisher Discriminant for the ϕK_S^0 analysis.

7.5.2 $B^0 \rightarrow \phi K_L^0$ Dataset

For the K_L^0 data sample, we prepared a set of five discriminating variables from signal Monte-Carlo and off-resonance data. We compared the training performance of a number of alternate training methods and for three different network sizes (one, two, and five hidden layer nodes). In order to compare the computing efficiencies of the alternate

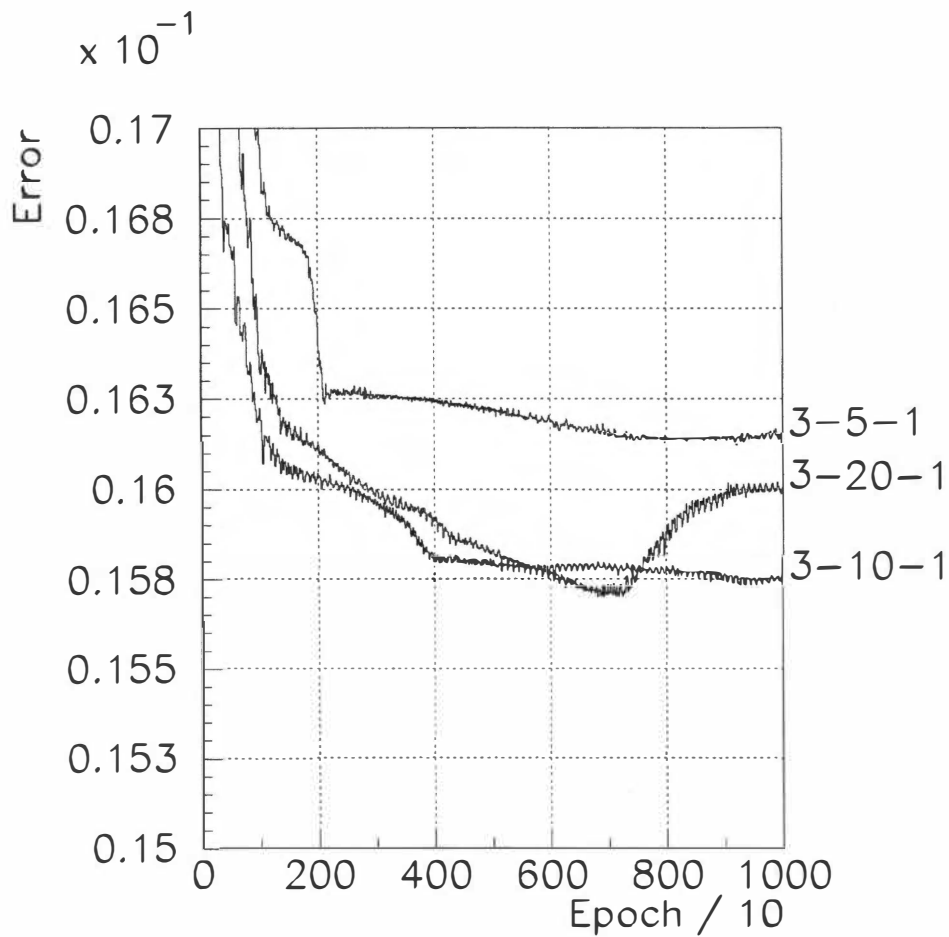


Figure 7.2: This plot shows the progress of the MLP network Error (Equation 7.2) as a function of training epoch for three different network sizes. There is an initial downward trend for all errors. However, larger networks can become over-trained and eventually result in an increased error after a number of epochs (as seen in the 3-20-1 network after epoch 700). The training should be halted when the error reaches a minimum.

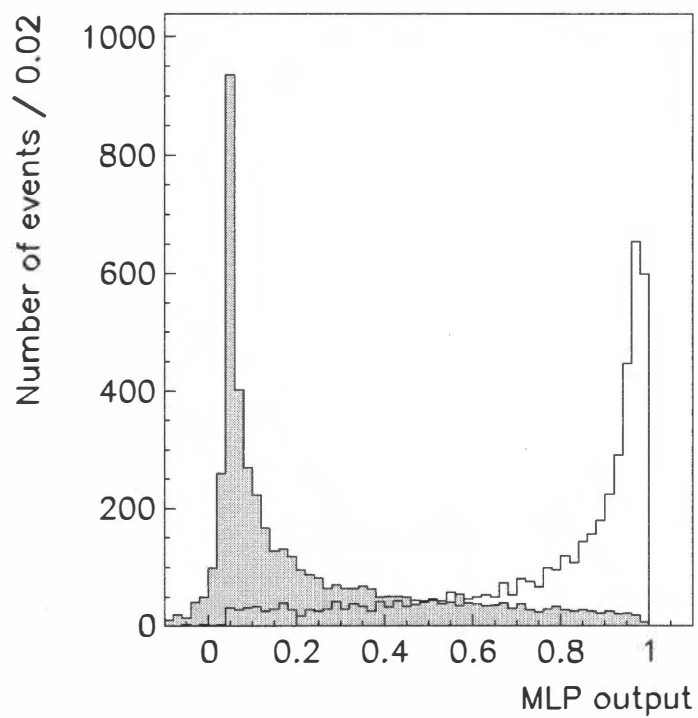


Figure 7.3: This shows the distribution of the MLP output for background (shaded) and signal (clear) events in an optimized background separation MLP consisting of four input nodes, four hidden layer nodes, and one output node.

Table 7.1: This is an investigation of the dependence of the final MLP output Error on network size with the objective of choosing a network which results in the lowest Error after a set number of training epochs (10,000). All networks used a single hidden layer. σ_{Error} was calculated from ten separate training attempts for each trial. For this data sample ($B^0 \rightarrow \phi K_S^0$, $K_S^0 \rightarrow \pi^0 \pi^0$), there appears to be little advantage in using a network having more hidden nodes than number of input variables.

Hidden layer nodes	Error _{min}	Error _{avg}	σ_{Error}
1	0.3625	0.3626	0.000078
2	0.3605	0.3616	0.00099
3	0.3599	0.3615	0.0012
4	0.3597	0.3611	0.0012
5	0.3592	0.3609	0.0012
6	0.3598	0.3607	0.0009
7	0.3594	0.3607	0.0011
8	0.3598	0.3607	0.0007
9	0.3593	0.3606	0.0011

training algorithms, we recorded the minimum number of epochs required to reach a minimized Error (Table 7.2).

From these results, we concluded that training method #6 was generally the most efficient but did not produce the most highly optimized network. The stochastic minimization algorithm (#1) produced acceptable error but was among the less efficient methods. It also performed better on larger network sizes. The MLP with five hidden layers produced a significant improvement over the Fisher discriminant and therefore was used for background suppression in the analysis of $B^0 \rightarrow \phi K_L^0$ [73].

7.6 Conclusion

The primary purpose of this investigation is the evaluation of the suitability of an MLP-based background separation method over the currently used Fisher Discriminant. We showed that the simple MLP network with no hidden layers and a single output is equivalent to the Fisher Discriminant and is thus a good indication of its performance on this data set. We conclude that a small improvement in continuum background suppression is possible through use of a more complex algorithm than the Fisher discriminant in the $B^0 \rightarrow \phi K_L^0$ analysis.

In particular, a small MLP network size and the stochastic minimization training algorithm is the optimal solution for the data sets we tested. This MLP consists of one hidden layer and a number of hidden layer nodes equal to the number of MLP inputs. More complex networks did not result in a sufficient increase in background separation performance to warrant the increased training and evaluation time required.

The MLP turns out to provide significant improvement for the ϕK_L^0 final state (which has high background due to its limited luminatic reconstruction).

Table 7.2: This is an investigation of the effect of training algorithm and MLP network size on training speed and final MLP Error. Each network was sufficiently trained to ensure convergence of the Error to a minimum. Training algorithms are detailed in Ref. [1].

MLP size	Training Alg.	N epochs	Error _{min}
$5 \rightarrow 1 \rightarrow 1$	1	5000	0.317
$5 \rightarrow 1 \rightarrow 1$	2	15000	0.3238
$5 \rightarrow 1 \rightarrow 1$	3	15000	0.3175
$5 \rightarrow 1 \rightarrow 1$	4	10000	0.3177
$5 \rightarrow 1 \rightarrow 1$	5	10000	0.3177
$5 \rightarrow 1 \rightarrow 1$	6	10000	0.3177
$5 \rightarrow 1 \rightarrow 1$	7	10000	0.3177
$5 \rightarrow 2 \rightarrow 1$	1	5000	0.315
$5 \rightarrow 2 \rightarrow 1$	2	10000	0.332
$5 \rightarrow 2 \rightarrow 1$	3	10000	0.318
$5 \rightarrow 2 \rightarrow 1$	4	1600	0.314
$5 \rightarrow 2 \rightarrow 1$	5	5000	0.316
$5 \rightarrow 2 \rightarrow 1$	6	400	0.312
$5 \rightarrow 2 \rightarrow 1$	7	100	0.318
$5 \rightarrow 5 \rightarrow 1$	1	6400	0.312
$5 \rightarrow 5 \rightarrow 1$	2	6400	0.330
$5 \rightarrow 5 \rightarrow 1$	3	6400	0.317
$5 \rightarrow 5 \rightarrow 1$	4	6400	0.311
$5 \rightarrow 5 \rightarrow 1$	5	6400	0.317
$5 \rightarrow 5 \rightarrow 1$	6	400	0.318
$5 \rightarrow 5 \rightarrow 1$	7	1600	0.333

Chapter 8

Analysis

The signal yield for the decay $B^0 \rightarrow \phi K_S^0$, $\phi \rightarrow K^+ K^-$, and $K_S^0 \rightarrow \pi^0 \pi^0 \rightarrow \gamma\gamma\gamma\gamma$ is extracted with an extended maximum likelihood fit, taking full advantage of the statistics with the knowledge of the distributions of the event variables. We also perform a cut-and-count analysis as a cross check on the signal yield.

8.1 Cut-and-Count Analysis

In a cut-and-count analysis, requirements for the event variable ranges, to select signal and suppress background, are optimized either for a branching fraction (BF) measurement (with a significant signal observation) or for an upper limit measurement.

To calculate the branching ratio (Equation 8.1), we need to know the total number of data events measured by the detector (227×10^6 events) and the total signal efficiency (divided into reconstruction, decay, and cut efficiencies). Reconstruction efficiency (ϵ_{rec}) is the ratio of signal events which pass the reconstruction of the decay channel as determined by signal Monte-Carlo. The cut efficiency (ϵ_{cut}) is the ratio of signal events which pass the cuts on the Fisher and Helicity variables. The decay efficiency (ϵ_{dec}) is the ratio between the branching fractions of the channel we wish to measure ($B^0 \rightarrow \phi K^0$) and the decay final state that we use to measure it ($\phi \rightarrow K^+ K^-$, $K^0 \rightarrow K_S^0$, $K_S^0 \rightarrow \pi^0 \pi^0$, $\pi^0 \rightarrow \gamma\gamma$). This ratio is the product of probabilities of each constituent decay. Given the number of signal events N_S we calculate the BF as follows:

$$BF(B^0 \rightarrow \phi K^0) = \frac{N_S}{\epsilon_{\text{rec}} \epsilon_{\text{cut}} \epsilon_{\text{dec}} N_{B\bar{B}}} \quad (8.1)$$

From the branching fraction, $BF(B^0 \rightarrow \phi K^0) = 8.6 \times 10^{-6}$ [74], and the sub-decays $BF(\phi \rightarrow K^+ K^-) = 0.491$, $BF(K^0 \rightarrow K_S^0) = 0.5$, and $BF(K_S^0 \rightarrow \pi^0 \pi^0) = 0.3$ [23],

we expect to observe about 20 signal events and, therefore, optimize our cuts with the expectation of a signal observation.

We define a signal region (signal box) and a background region (sideband) in the ΔE versus m_{ES} plane. The input samples are on-resonance and off-resonance data (for background) and Monte-Carlo (MC) simulated signal events. We perform the optimization with a blinded procedure in which the data events in the signal region are explicitly excluded [75] until signal yield and all background contributions are determined, and the systematic uncertainties in the BF are estimated.

8.1.1 Optimized Cuts

There are a number of criteria (“Figures of Merit”) which can be used to determine optimal requirements on the event variables. Given the number of signal events (N_s) and number of background events (N_b) which pass a given criterion, commonly used Figures of Merit are the signal purity $N_s/(N_s + N_b)$ and the signal significance $N_s/\sqrt{N_s + N_b}$. For the purposes of a BF measurement we chose the signal significance Figure of Merit.

We use the four event variables (m_{ES} , ΔE , Fisher, and Helicity), and optimize criteria for the Fisher Discriminant and the Helicity variables, scanning the full 2-dimensional parameter space. Optimum requirements are $\mathcal{F} > -1.175$ (suppressing 72% of background and 14% of signal) and $|\cos \theta_H| > 0.361$ (suppressing 32% of background and 5% of signal). The combined requirement suppresses 81% of background and 18% of signal. Scan projections of the estimated signal significance versus event variable requirement are shown in Figure 8.1 for the Fisher Discriminant and Figure 8.2 for the Helicity variable.

8.1.2 Counting Analysis

We define a 2σ wide signal region in the ΔE versus m_{ES} plane around the signal peak: the interval $(5.27, 5.29) \text{ GeV}/c^2$ for m_{ES} and $(-0.12, 0.056) \text{ GeV}$ for ΔE (Figure 8.3). The resolutions (σ) are obtained from signal Monte-Carlo. We also define a region dominated by background known as the Grand Sideband (GSB) as $m_{\text{ES}} < 5.265 \text{ GeV}/c^2$. We use the GSB region in addition to the portion of ΔE outside of the signal region to estimate the number of background events in the signal region. Regions for the GSB, ΔE sidebands, and the signal box are listed in Table 8.1 and presented in Figure 8.4.

We assume that the ΔE distribution in regions outside the signal box are representative of background events in the signal region. We confirm this with Monte-Carlo studies of continuum background, using on-resonance samples, to show that contributions from

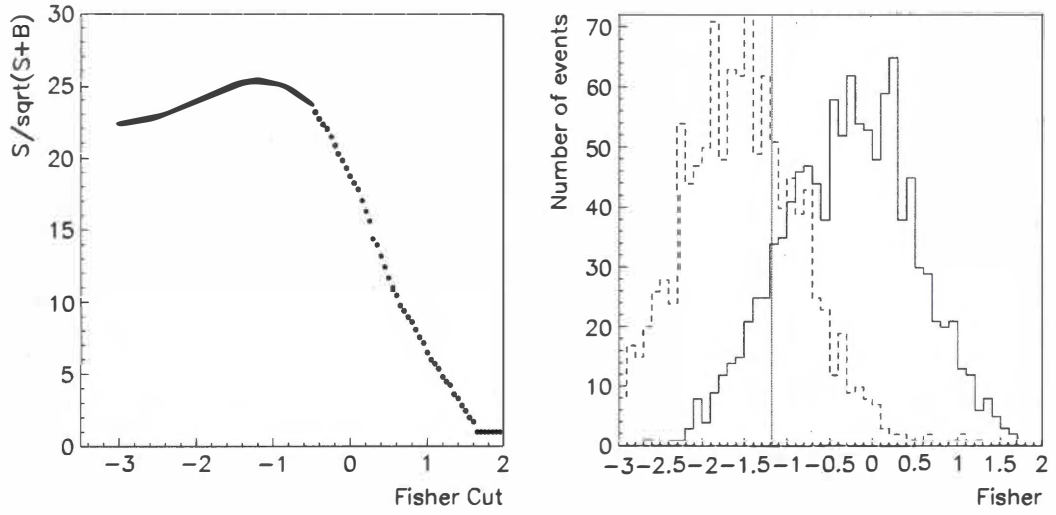


Figure 8.1: The plot on the left shows the signal significance versus range cutoff for the Fisher event variable. The right plot shows the location of the optimized cut (vertical line) (maximum signal significance) on the variable's distributions for signal MC (solid) and sideband on-resonance background (dashed) samples.

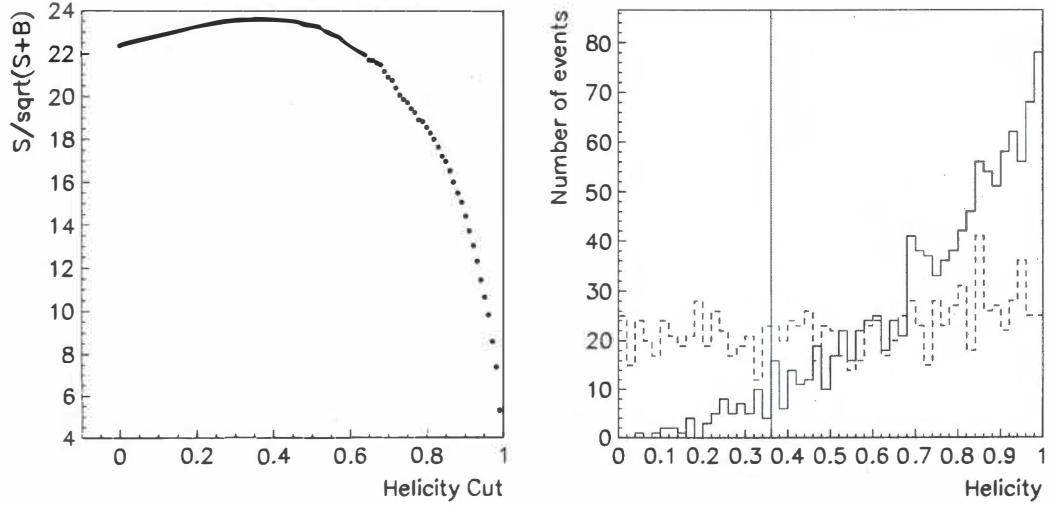


Figure 8.2: The plot on the left shows the signal significance versus cutoff for the Helicity event variable ($|\cos \theta_H|$). The right plot shows the location of the optimized cut (vertical line) (maximum signal significance) on the variable's distributions for signal MC (solid) and sideband on-resonance background (dashed) samples.

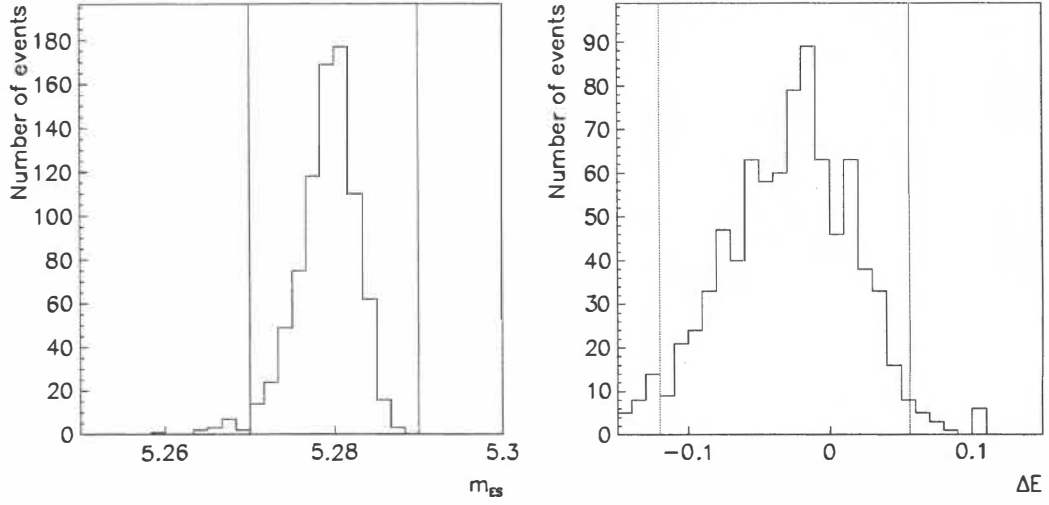


Figure 8.3: These plots show the distributions of signal MC events in the signal box interval for m_{ES} (left) and ΔE (right).

Table 8.1: The three regions used for the cut-and-count analysis: ΔE sideband, Grand Sideband, and the signal region

Region	Range
ΔE sideband	$\Delta E > 0.056 \text{ GeV}$ $\Delta E < -0.12 \text{ GeV}$
Grand Sideband	$m_{ES} < 5.265 \text{ GeV}/c^2$
Signal Region	$m_{ES} > 5.27 \text{ GeV}/c^2$ $-0.12 \text{ GeV} < \Delta E < 0.056 \text{ GeV}$

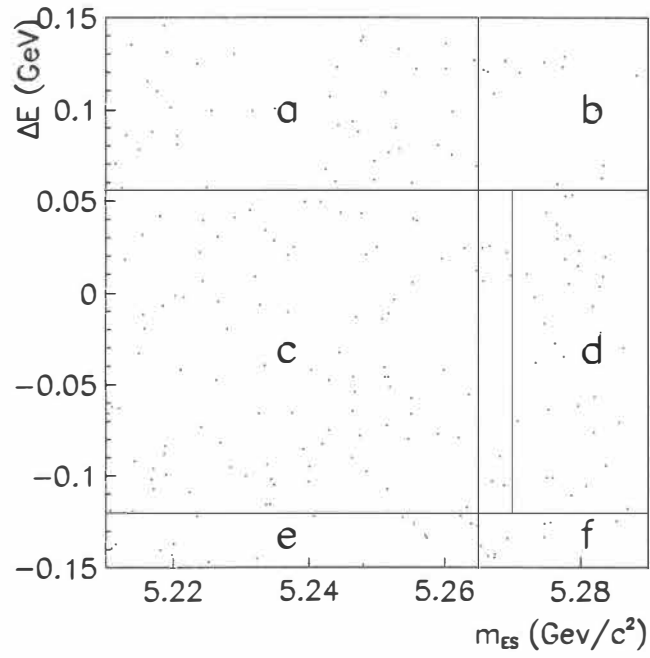


Figure 8.4: The regions in ΔE and m_{ES} listed in Table 8.1 are identified on the plot: ΔE sideband (a, b, e, and f), Grand Sideband (a, c and e), and the signal region (d).

Table 8.2: These correlation coefficients between the event variables (from signal Monte-Carlo and sideband on-resonance background events) include the measured proper decay time difference Δt between the two B mesons in the event and the error $\sigma_{\Delta t}$. The variables will be explained in detail in Section 8.3.2.

	ΔE	Fisher	$ \cos \theta_H $	Δt	$\sigma_{\Delta t}$
Signal Monte-Carlo					
m_{ES}	0.0617	0.0051	0.0007	0.0020	-0.0104
ΔE		-0.0056	-0.0026	-0.0049	-0.0493
Fisher			-0.0022	0.0014	0.0105
$ \cos \theta_H $				-0.0012	0.0095
Δt					-0.0194
Sideband Data					
m_{ES}	-0.0220	0.0227	-0.0561	0.0291	-0.0579
ΔE		-0.0028	0.0048	0.0296	-0.0028
Fisher			0.0485	-0.0458	0.0416
$ \cos \theta_H $				-0.0327	0.0683
Δt					0.1146

B background are negligible (see Chapter 5.2). We also compare background variable shapes in on and off-resonance data. Furthermore, the correlation coefficients among the event variables in Monte-Carlo signal and sideband data (Table 8.2) are measured to be small enough to justify that the variables can be treated independently. We parameterize the ΔE distribution with a linear function $f(x) = N(1 + m \cdot x)$ (where m is the slope and N is the number of events) and fit it to data events in the GSB region (Figure 8.5). The slope is the only relevant parameter for calculation of the ratio of background events between the signal region and the ΔE sidebands.

We fit the GSB events to a phase-space (Argus shape) function [76] shown in Figure 8.6:

$$\text{Argus}(x) = N(x/x_1)\sqrt{1 - (x/x_1)^2}e^{\zeta(1-(x/x_1)^2)}. \quad (8.2)$$

With $x = m_{ES}$, the parameters for the Argus function are the shape parameter ζ and the end point x_1 (the largest non-zero value the function can take). For on-resonance data, we fix the end point to $x_1 = 5.29 \text{ GeV}/c^2$ and fit for the shape parameter in the ΔE sidebands. We cross-check the fit for the shape parameter with events in the GSB. With the shape parameter known, we integrate the Argus function in both the sideband and signal regions. The ratio of the integrals is the scale factor for the number of background events in the signal box (Figure 8.7).

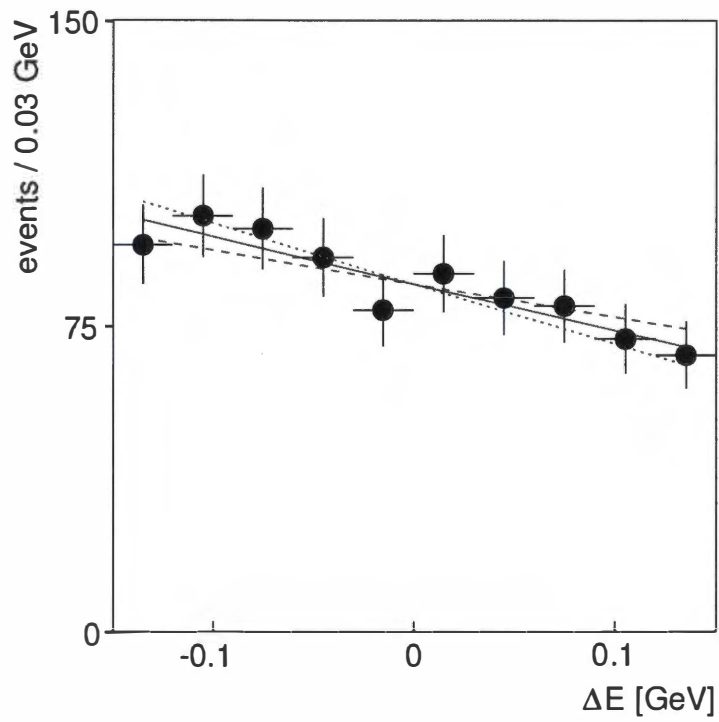


Figure 8.5: The ΔE distribution in data from the GSB (dots) with the fit to a linear function superimposed and dotted lines indicating the 1σ range of uncertainty in the fitted slope value

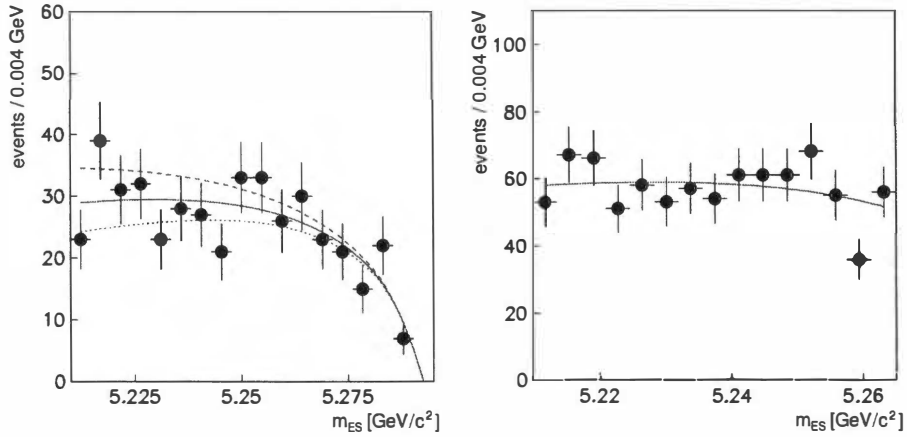


Figure 8.6: The m_{ES} distribution is parameterized with an “Argus” shape function (Equation 8.2) and is fit to data events in the ΔE sidebands (ranges $\Delta E < -0.012$ GeV and $\Delta E > 0.056$ GeV) (left). The dotted lines show the $\pm 1\sigma$ ranges of uncertainty in the Argus shape parameter. We fit m_{ES} to data in the Grand Sideband (GSB) region ($m_{ES} < 5.265$ GeV/c) (right) as a cross check on the shape parameter. Both parameter sets are in good agreement.

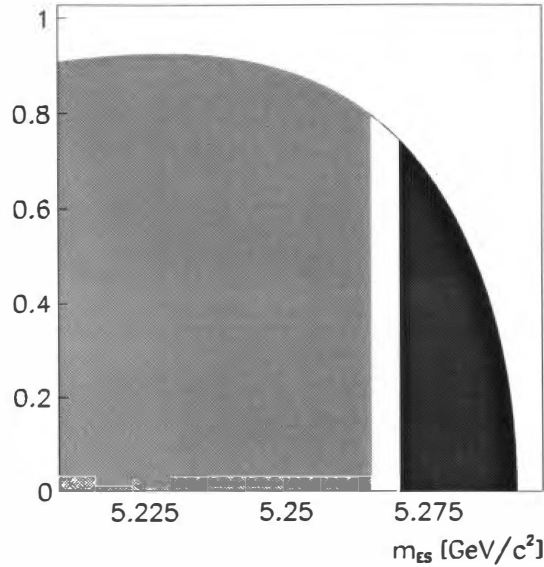


Figure 8.7: We calculate the number of background events expected in the m_{ES} signal region (right shaded area) by scaling the number of events counted in the GSB (left shaded area) by the ratio of areas under the fitted Argus shape function.

The total scale factor between the signal region (SR) and GSB is the product of the scaling factors for ΔE and m_{ES} . We estimate the number of background events in the SR by counting and scaling the events in the GSB and then subtract the background events from the observed number of data events in the SR. We calculate the BF to according to Equation 8.1 to be

$$BF(B^0 \rightarrow \phi K^0) = (5.21 \pm 2.46) \times 10^{-6}. \quad (8.3)$$

Quantitative results for the counting analysis are presented in Table 8.3.

8.1.3 B Background

The introduction of signal-like background from B^+B^- and $B^0\bar{B}^0$ events has not been explicitly taken into account in this analysis. From our analysis of $B^0\bar{B}^0$ generic Monte-Carlo data, we found that no such background events pass the reconstruction requirements. From B^+B^- background study we expect 2.3 background (primarily from f^0K^0 events). However, due to energy leakage, these events appear lower in ΔE than our signal and are unlikely to appear in the signal box.

8.1.4 Box-counting Cross-check

We perform another consistency check with a pure box-counting method. Figure 8.8 shows the 6 regions of interest in our ΔE versus m_{ES} plot. To a good approximation, the background events are linearly distributed in ΔE . This allows us to use the number of events in each region, N_i , (labeled $i = \text{“a”}$ through “f” in Figure 8.8) to calculate the number of background events in the signal region “d” , N_d :

$$N_d = (N_{d_1} + N_{d_2})/2 \quad \text{where} \quad N_{d_1} = \frac{N_c N_f}{N_e} \quad \text{and} \quad N_{d_2} = \frac{N_c N_b}{N_a}. \quad (8.4)$$

Our final event counts for the 5 known regions are as follows: $N_a = 46$, $N_b = 10$, $N_c = 97$, $N_e = 20$, and $N_f = 4$. We find $N_d = 34$ events in the signal region, $N_{\text{bkg}} = 20.2 \pm 6.6$ background events, and therefore $N_{\text{sig}} = 13.8 \pm 8.8$ signal events. With the reconstruction, cut, and decay efficiencies as well as the total number of $B\bar{B}$ events, the BF of $B^0 \rightarrow \phi K^0$ is

$$BF(B^0 \rightarrow \phi K^0) = (5.8 \pm 3.7) \times 10^{-6}, \quad (8.5)$$

which is in good agreement with our previous estimate.

Table 8.3: Results from the Cut and Count cross-check analysis for signal yield

Item	Value
ΔE slope	-1.35 ± 0.39
ΔE scale factor	0.6123 ± 0.0073
from ΔE sideband	
m_{ES} shape parameter	-21.4 ± 6.2
m_{ES} scale factor	0.217 ± 0.02
from GSB (cross-check, unused)	
m_{ES} shape parameter	-20.4 ± 6.8
Total scale factor (GSB $\rightarrow$ Signal Region)	0.1332 ± 0.0123
Event Counting	
GSB background	163 ± 12.8 events
Signal Region background	21.7 ± 2.6 events
Data in Signal Region	34 ± 5.8 events
Observed signal	12.3 ± 5.8 events
$N_{B\bar{B}}$	227×10^6 events
Reconstruction efficiency (ϵ_{rec})	$(17 \pm 0.3)\%$
Signal (background) cut efficiency	
Fisher > -1.175	86% (28%)
Helicity > 0.361	95% (68%)
combined ϵ_{cut}	82% (19%)
Decay efficiency [23]	
$\phi \rightarrow K^+ K^-$	$(49.2 \pm 0.6)\%$
$K^0 \rightarrow K_s^0$	50%
$K_s^0 \rightarrow \pi^0 \pi^0$	$(31.05 \pm 1.4)\%$
$\pi^0 \rightarrow \gamma\gamma$	$(98.798 \pm 0.032)^2\%$
combined ϵ_{dec}	$(7.46 \pm 0.0971)\%$
$BF(B^0 \rightarrow \phi K^0)$	$(5.21 \pm 2.46) \times 10^{-6}$

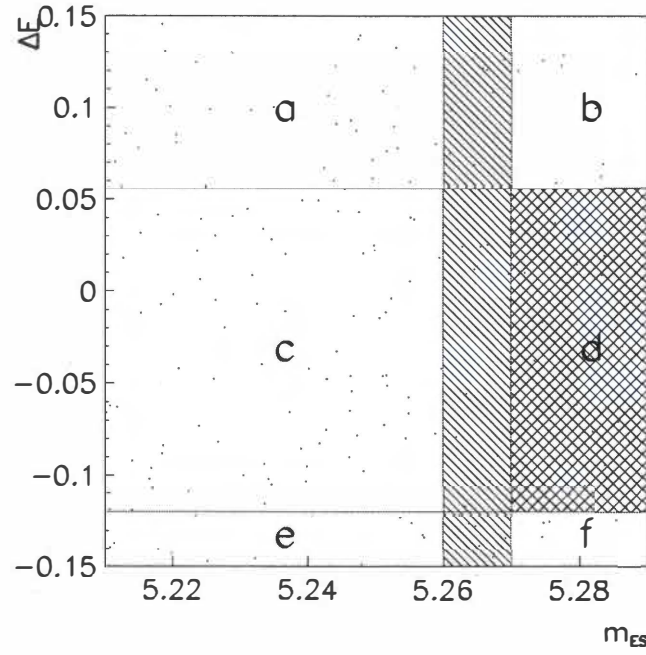


Figure 8.8: This plot of post-cut experimental data shows the signal region (cross-hatched). The region $5.26 \text{ GeV}/c^2 < m_{\text{ES}} < 5.27 \text{ GeV}/c^2$ (diagonal lines) between the sideband region and the signal region is ignored for background estimates.

8.2 Maximum Likelihood Fit to Event Yields

We use a maximum likelihood estimation to deduce the number of signal and background events. In our model, event variables distribute according to certain functional forms established from Monte Carlo simulations and data control samples for signal and from Monte Carlo simulations, side-band data, and off-resonance data for background. The normalized functions are used as probability density functions (PDF). For each event in the data set, the probability that this event occurs is the product of the PDFs. The complete likelihood is defined as

$$L = \exp\left(-\sum_{i=1}^M N_i\right) \prod_{k=1}^N \left(\sum_{l=1}^M N_l P_l(\vec{x}(\vec{\alpha}))\right)_k, \quad (8.6)$$

where the $P_l(\vec{\alpha}, \vec{x})$ are the product of PDFs for each of the fit input variables $\vec{x}$ (characterizing signal or background). Their parameters $\vec{\alpha}$ are extracted beforehand and fixed in the fit. The P_l are summed over two terms, one for signal and one for background ($M = 2$), with the N_l being the corresponding event yields. The first term takes the Poisson fluctuations in the total number of events into account. This likelihood is called extended maximum likelihood [38]. The N_l are the only free parameters in the likelihood maximization.

To construct the likelihood function and perform the maximization on the reconstructed data, we use the *RooFit* [50] package, developed for the *BABAR* experiment. It minimizes the quantity $-\ln(L)$ using the MINUIT fitting package [77]. The corresponding χ^2 is defined as

$$\chi^2 = -2 \cdot \ln \left(\frac{L}{L_{max}} \right), \quad (8.7)$$

with L_{max} the maximum likelihood value. The statistical significance S of the fitted signal yield N_{sig} is

$$S = \sqrt{-2 \cdot \ln \left(\frac{L(N_{sig} = 0)}{L_{max}} \right)}. \quad (8.8)$$

Our maximum likelihood fit is based on the four event variables that we established previously: m_{ES} , ΔE , Fisher discriminant, and $|\cos \theta_H|$. The maximum likelihood fit to event yields is done in such a way that it can be extended to a *CP* asymmetry fit.

8.2.1 Event Variable Parameterization

We model each event variable distribution with phenomenological functions. Whenever possible, the PDF form is based on known physics processes or statistical understanding. We divide the signal Monte-Carlo set into well reconstructed and self cross-feed (SXF) events. For each variable we construct PDFs on the two sets separately and add them with a relative fraction to form a total signal PDF: $\text{PDF}_{\text{tot}} = (1 - f) \text{PDF}_1 + f \text{PDF}_2$. The f is the relative fraction which is fixed in our fit to $f = 0.85$ (as determined from our photon-association study in Section 6.2.2). For the continuum background, the event variable distributions in the two categories “well-reconstructed” and “SXF” are identical. The B -decay background and event variable differences between the two categories are insignificant.

The m_{ES} PDF for well-reconstructed signal events is described by a sum of two Gaussians ($N = 1$; double-Gaussian):

$$\sum_{i=0}^{N-1} f_i G_i(x) + (1 - f_n) G_N(x), \quad (8.9)$$

with $x = m_{\text{ES}}$, m_i the mean and σ_i the standard deviation for Gaussian i , which can be written as:

$$G_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-m_i)^2}{2\sigma_i^2}}. \quad (8.10)$$

The m_{ES} PDF for SXF events is a sum of three Gaussians ($N = 2$). The (continuum) background m_{ES} distribution is modeled with an Argus shape function [76], with the endpoint fixed at $5.29 \text{ GeV}/c^2$.

For well reconstructed events, the ΔE PDF is a bifurcated Gaussian plus a polynomial of degree 1 (linear). In the bifurcated Gaussian, the standard deviation σ is parameterized as a step-function:

$$\sigma(x) = \begin{cases} \sigma_1 & x < \mu \\ \sigma_2 & x \geq \mu \end{cases}. \quad (8.11)$$

This results in a Gaussian distribution having a left tail with a length defined by σ_1 and a right tail with a length defined by σ_2 . For SXF events, the PDF is a sum of a symmetric Gaussian and a linear term. The background ΔE PDF is a linear function.

The Fisher discriminant is parameterized with a double Gaussian for both well-reconstructed and SXF events and with a single Gaussian for background.

The cosine of the Helicity angle for the signal is expected to distribute according to a $\cos^2 \theta_H$ function. Hence, we parameterize it with a second-order polynomial. The

background is expected to distribute nearly uniformly. We parameterize it with a linear function.

The parameterization and parameters are summarized in Table 8.4. Figure 8.9 shows these PDFs fit to well-reconstructed signal MC events, and Figure 8.10 shows the PDFs for SXF signal MC events.

8.2.2 Fit Stability Analysis

We test the stability, variation, and potential bias of the yield fit result with fits to data sets generated from the model PDFs (*toy MC*). Each sample statistically reflects the expected signal and background composition and the distribution of the event variables in the data by randomly choosing events based on the event probabilities. Since these toy events are generated from independent PDFs, there are no correlations between the different variables, as is possible in real data. For our analysis, correlations between the four event variables are negligible (see Table 8.2). Hence, toy MC is appropriate to study the fit characteristics. The distribution of the fitted signal yield from 2000 toy data sets with expected 19 signal and 1100 background events is displayed in Figure 8.11 (left). The width of the distribution is about 6.4. The distribution in Figure 8.11 is the pull (or normalized Gaussian) defined as:

$$\text{pull} = \frac{N - N_0}{\sigma}, \quad (8.12)$$

where $N_0 = 19$ is the mean and σ the event-by-event error on the yield N . This quantity is expected to be centered at zero and to have a width of 1.0. The mean of the pull distribution in Figure 8.11 is marginally different from 0.0 and the width is consistent with 1.0.

As a cross check, we test the influence of residual correlations among the signal variables by randomly picking 19 signal events from a set of fully reconstructed MC events and mixing them together with background events as generated from PDFs. In 100 such *mock* fits we did not observe any bias.

We also provide a comparison of the event-by-event likelihood values from toy sets and from real data. They are plotted in Figure 8.12 with a logarithmic event scale. The compatibility between the likelihood distributions for toy MC (solid histogram) and data (points with error bars) is an indication that the likelihood treatment of data is well modeled by the toy MC analysis.

From the toy MC analysis, we can conclude that our PDFs are an unbiased representation of the variable distributions in data and that the expected statistical error of the signal yield is 6.4 events.

Table 8.4: Parameterization of the event variable PDFs for the final state $B^0 \rightarrow \phi K_s(\pi^0\pi^0)$ (kinematic quantities in GeV)

Variable	Signal	SXF Signal	Background
m_{ES}	Double Gaussian $m_0 = 5.280 \pm 0.00028$ $\sigma_0 = 0.0028 \pm 0.00021$ $m_1 = 5.275 \pm 0.00025$ $\sigma_1 = 0.0045 \pm 0.0008$ $f = 0.85 \pm 0.011$	Triple Gaussian $m_0 = 5.280 \pm 0.00014$ $\sigma_0 = 0.0030 \pm 0.000078$ $m_1 = 5.265 \pm 0.0047$ $\sigma_1 = 0.0081 \pm 0.0020$ $m_2 = 5.275 \pm 0.00064$ $\sigma_2 = 0.0045 \pm 0.0003$ $f_0 = 0.666 \pm 0.0416$ $f_1 = 0.0270 \pm 0.0173$	Argus Function $\xi = -22.0 \pm 3.85$
ΔE	Bifurcated Gaussian + Polynomial ₁ $m_0 = -0.0092 \pm 0.0006$ $\sigma_L = 0.0291 \pm 0.0004$ $\sigma_R = 0.0519 \pm 0.0007$ slope = -6.06 ± 0.13 $f = 0.0989 \pm 0.0059$	Gaussian + Polynomial ₁ $m_0 = -0.00384 \pm 0.0026$ $\sigma = 0.00511 \pm 0.0029$ slope = -6.04 ± 0.144 $f = 0.461 \pm 0.0414$	Polynomial ₁ slope = -1.33 ± 0.34
Fisher	Double Gaussian $m_0 = -0.701 \pm 0.073$ $\sigma_0 = 0.676 \pm 0.021$ $m_1 = 0.210 \pm 0.050$ $\sigma_1 = 0.596 \pm 0.015$ $f = 0.458 \pm 0.069$	Double Gaussian $m_0 = -0.719 \pm 0.13$ $\sigma_0 = 0.680 \pm 0.043$ $m_1 = 0.225 \pm 0.22$ $\sigma_1 = 0.574 \pm 0.040$ $f = 0.350 \pm 0.14$	Gaussian $m_0 = -1.57 \pm 0.022$ $\sigma = 0.706 \pm 0.016$
$ \cos \theta_H $	Polynomial ₂ linear = -25.6 ± 16.3 quad = 1258.0 ± 1200	Polynomial ₂ linear = -18.0 ± 8.63 quad = 523.5 ± 297.3	Polynomial ₁ slope = 0.324 ± 0.136

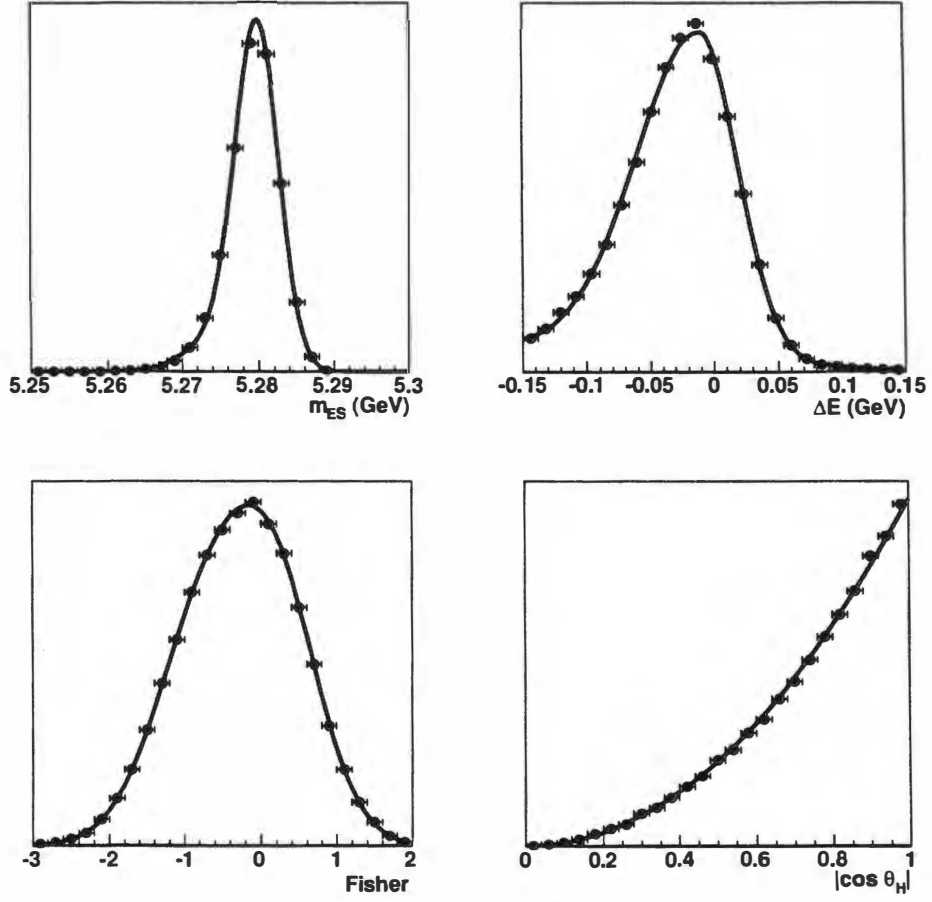
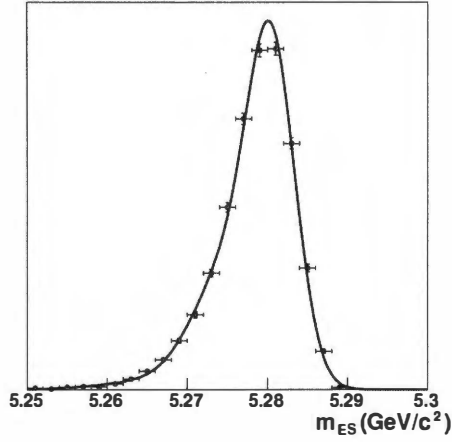
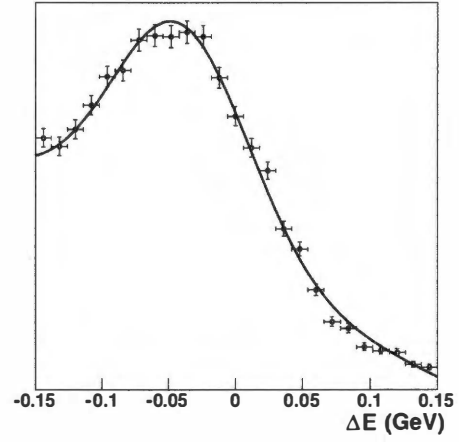


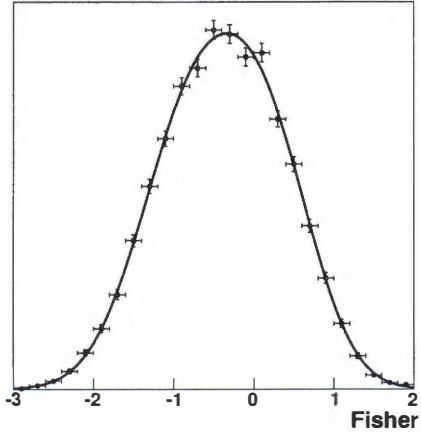
Figure 8.9: The Four event variable parameterizations for well-reconstructed signal events fitted on signal MC data with an arbitrary normalization



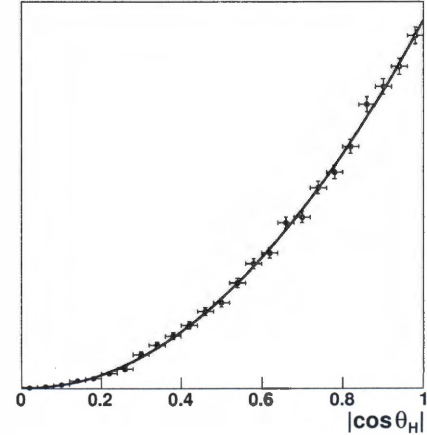
(a)



(b)



(c)



(d)

Figure 8.10: The parameterizations of the four event variables for SXF signal events fitted on SXF signal MC with arbitrary normalization

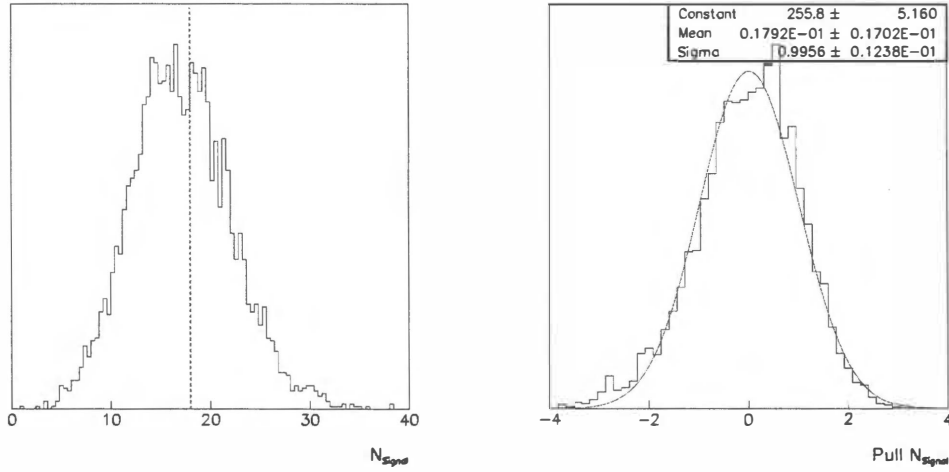


Figure 8.11: The left figure shows the distribution of our result N_{sig} from fits to 2000 toy data set. The dotted line represents our result. The right plot shows the “pull” distribution of the signal yield from these fits. The variable “pull” is explained in the text. Both histograms have an arbitrary normalization.

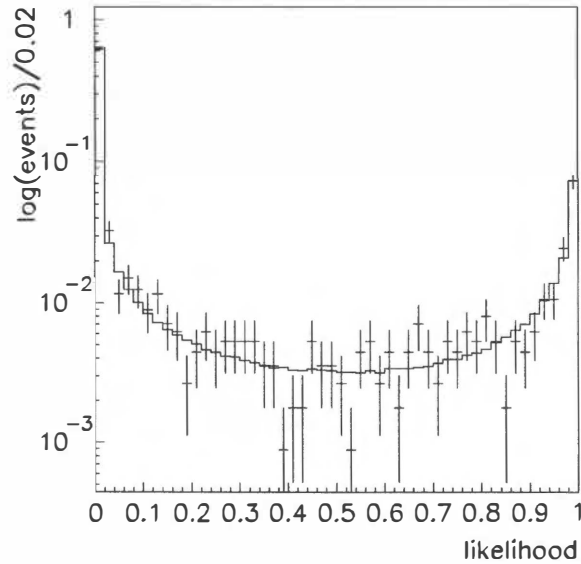


Figure 8.12: Histograms of per-event $\log(\text{likelihood})$ for toy MC (solid) and experiment data (points with error bars) show that the likelihood treatment is similar for both sets of events. Event yields have been normalized and are shown on a logarithmic scale.

8.2.3 Systematic Uncertainties

In the following we estimate the dominant systematic uncertainties for the signal yield.

Fit Parameters

Our PDFs for signal are modeled on signal MC and the parameters are then fixed in the final fit. The uncertainty in these parameters in the fit to MC events is negligible due to the large amount of available signal MC. The agreement in event variable distributions in MC and data is tested with data control samples with larger sample size such as $B^0 \rightarrow J/\psi K_S, K_S \rightarrow \pi^0 \pi^0$ and inclusive $K_S \rightarrow \pi^0 \pi^0$ [78]; the differences do not contribute significantly. The systematic uncertainties for the PDF parameters is estimated by varying each parameter independently in the fit by plus and minus one standard deviation while fixing all other parameters except the signal and background yield. The corresponding shift in N_{sig} is taken as the systematic error of the respective parameter. The uncertainties from all fixed parameters in the fit are added in quadrature to calculate the total systematic error (see Table 8.5).

Reconstruction Efficiency of π^0

There is a 3% systematic uncertainty in the efficiency associated with each π^0 due to corrections of the photon energies. This was established by the neutral hadron analysis working group of *BABAR* from large data control samples [79].

SXF Fraction

The fraction of SXF signal events is determined from signal MC, but the error on this value due to differences between MC and data is difficult to measure. The recommended procedure in *BABAR* [80] is to study the effect on the yield by a $\pm 10\%$ variation in the fraction. We modify the SXF fraction by $\pm 10\%$ and take the resulting change in N_{sig} as the systematic error.

B^0 Counting

There is an error associated with the counting of $B\bar{B}$ events from the Pep-II luminosity of $\pm 1.1\%$ [81].

Total

Systematic uncertainties from all sources are added in quadrature. They are presented in Table 8.6.

Table 8.5: The systematic error due to signal PDF parameters is evaluated by calculating the relative shift of the number of signal events N_{sig} due to a shift in each PDF parameter by ± 1 standard deviation. The component errors were added in quadrature for the total PDF parameter systematic error.

Parameter	Value	Error on N_{sig}
Well-reconstructed Signal PDF		
m_{ES} G. mean ₁	5.28	$\pm 1.9 \times 10^{-5}$ 0.06%
m_{ES} G. σ_1	2.82×10^{-3}	$\pm 1.4 \times 10^{-5}$ 0.16%
m_{ES} G. mean ₂	5.275	$\pm 1.6 \times 10^{-4}$ 0.12%
m_{ES} G. σ_2	4.46×10^{-3}	$\pm 5.5 \times 10^{-5}$ 0.06%
m_{ES} G _{1,2} ratio	0.865	$\pm 6.7 \times 10^{-3}$ 0.16%
ΔE G. mean	-9.28×10^{-3}	$\pm 4.2 \times 10^{-4}$ 0.23%
ΔE G. σ_1	5.11×10^{-2}	$\pm 4.7 \times 10^{-4}$ 0.07%
ΔE G. σ_2	2.87×10^{-2}	$\pm 2.8 \times 10^{-4}$ 0.24%
ΔE slope	-6.208	$\pm 9.1 \times 10^{-2}$ 0.06%
ΔE weight	0.109	$\pm 4.3 \times 10^{-3}$ 0.15%
$\mathcal{F}$ G. mean ₁	-0.734	$\pm 4.4 \times 10^{-2}$ 0.95%
$\mathcal{F}$ G. σ_1	0.665	$\pm 1.3 \times 10^{-2}$ 0.27%
$\mathcal{F}$ G. mean ₂	0.198	$\pm 3.0 \times 10^{-2}$ 0.92%
$\mathcal{F}$ G. σ_2	0.595	$\pm 9.0 \times 10^{-3}$ 0.14%
$\mathcal{F}$ G _{1,2} ratio	0.484	$\pm 4.0 \times 10^{-2}$ 2.21%
$\mathcal{H}$ linear term	-20.759	± 5.8 0.11%
$\mathcal{H}$ quadratic term	1.98×10^3	$\pm 1.2 \times 10^3$ 0.18%
SXF Signal PDF		
m_{ES} G. mean ₁	5.281	$\pm 1.2 \times 10^{-4}$ 0.10%
m_{ES} G. σ_1	3.00×10^{-3}	$\pm 6.1 \times 10^{-5}$ 0.11%
m_{ES} G. mean ₂	5.267	$\pm 1.5 \times 10^{-3}$ 0.07%
m_{ES} G. σ_2	7.17×10^{-3}	$\pm 5.0 \times 10^{-4}$ 0.01%
m_{ES} G. mean ₃	5.275	$\pm 4.4 \times 10^{-4}$ 0.15%
m_{ES} G. σ_3	4.33×10^{-3}	$\pm 1.7 \times 10^{-4}$ 0.08%
m_{ES} G _{1,2} ratio	0.633	$\pm 3.3 \times 10^{-2}$ 0.10%
m_{ES} G _{1,3} ratio	4.19×10^{-2}	$\pm 1.2 \times 10^{-2}$ 0.14%
ΔE mean	-4.09×10^{-2}	$\pm 1.9 \times 10^{-3}$ 0.04%
ΔE σ	5.13×10^{-2}	$\pm 2.2 \times 10^{-3}$ 0.17%
ΔE linear term	-5.901	± 0.11 0.07%
ΔE component ratio	0.463	$\pm 3.1 \times 10^{-2}$ 0.17%
$\mathcal{F}$ G. mean ₁	-0.717	$\pm 1.4 \times 10^{-2}$ 0.07%
$\mathcal{F}$ G. σ_1	0.676	$\pm 1.0 \times 10^{-2}$ 0.05%
$\mathcal{F}$ G. mean ₂	0.229	$\pm 2.1 \times 10^{-2}$ 0.09%
$\mathcal{F}$ G. σ_2	0.569	$\pm 1.3 \times 10^{-2}$ 0.04%
$\mathcal{F}$ G _{1,2} ratio	0.359	$\pm 1.1 \times 10^{-2}$ 0.12%
$\mathcal{H}$ linear term	-10.017	± 3.6 0.07%
$\mathcal{H}$ quadratic term	3.56×10^2	$\pm 1.5 \times 10^2$ 0.03%
Background m_{ES} endpoint	5.29	$\pm 3.8 \times 10^{-4}$ 2.6%
Total fit parameter systematics		$\pm 3.1\%$

Table 8.6: Estimate of the total yield fit systematic error sources

Source	Error
PDF parameters (Table 8.5)	3.1%
π^0 reconstruction efficiency	4.2%
SXF fraction	3.6%
B counting	1.1%
total systematic error	6.4%

8.2.4 Result

The final extended maximum likelihood fit extracts signal and background yields together with all background parameters except the m_{ES} endpoint (see Table 8.7). The fit with 1129 data events results in

$$N_{\text{sig}} = 19.6^{+6.9}_{-5.8}. \quad (8.13)$$

This corresponds to a $B^0 \rightarrow \phi K^0$ branching fraction (Equation 8.1) of

$$BF(B^0 \rightarrow \phi K^0) = (6.6^{+2.4}_{-2.0}(\text{stat}) \pm 0.42(\text{syst})) \times 10^{-6}. \quad (8.14)$$

The statistical significance of the measurement is $\chi = \sqrt{-2 \ln L_0/L_{\text{max}}} = 5.6$ where L_0 is the likelihood evaluated at $N_{\text{sig}} = 0$. We plot $-\ln L/L_{\text{max}}$ versus N_{sig} in Figure 8.13. Figure 8.14 shows the distributions of the four event variables with the result of the maximum likelihood fit superimposed (solid line). For demonstration purpose the signal in those plots is enhanced by a requirement on the event likelihood: the PDF for the respective variable is removed from the likelihood calculation and then a criterion is applied to the ratio of the probabilities (or partial likelihoods) for signal (L_S) and background (L_B):

$$r < \frac{L_S}{L_S + L_B}, \quad (8.15)$$

with r chosen to suppress background. The efficiency for signal and background is evaluated with MC events. The fit result is scaled accordingly.

Table 8.7: Results from the branching fraction analysis

Item	Value
Background Parameters	
ΔE slope	-1.33 ± 0.34
$\mathcal{F}$ mean	-1.57 ± 0.02
$\mathcal{F}\sigma$	0.71 ± 0.02
$\mathcal{H}$ slope	0.32 ± 0.14
N_{bkg}	1110 ± 34
N_{sig}	$19.6^{+6.9}_{-5.8}$
$N(B\bar{B})$	227×10^6 events
Reconstruction efficiency (ϵ_{rec})	$(17 \pm 0.3)\%$
Decay efficiency [23]	
$\phi \rightarrow K^+ K^-$	$(49.2 \pm 0.6)\%$
$K^0 \rightarrow K_S^0$	50%
$K_S^0 \rightarrow \pi^0 \pi^0$	$(31.05 \pm 1.4)\%$
$\pi^0 \rightarrow \gamma\gamma$	$(98.798 \pm 0.032)^2\%$
combined ϵ_{dec}	$(7.46 \pm 0.0971)\%$
$BF(B^0 \rightarrow \phi K^0)$	$(6.6^{+2.4}_{-2.0}(\text{stat}) \pm 0.42(\text{syst})) \times 10^{-6}$

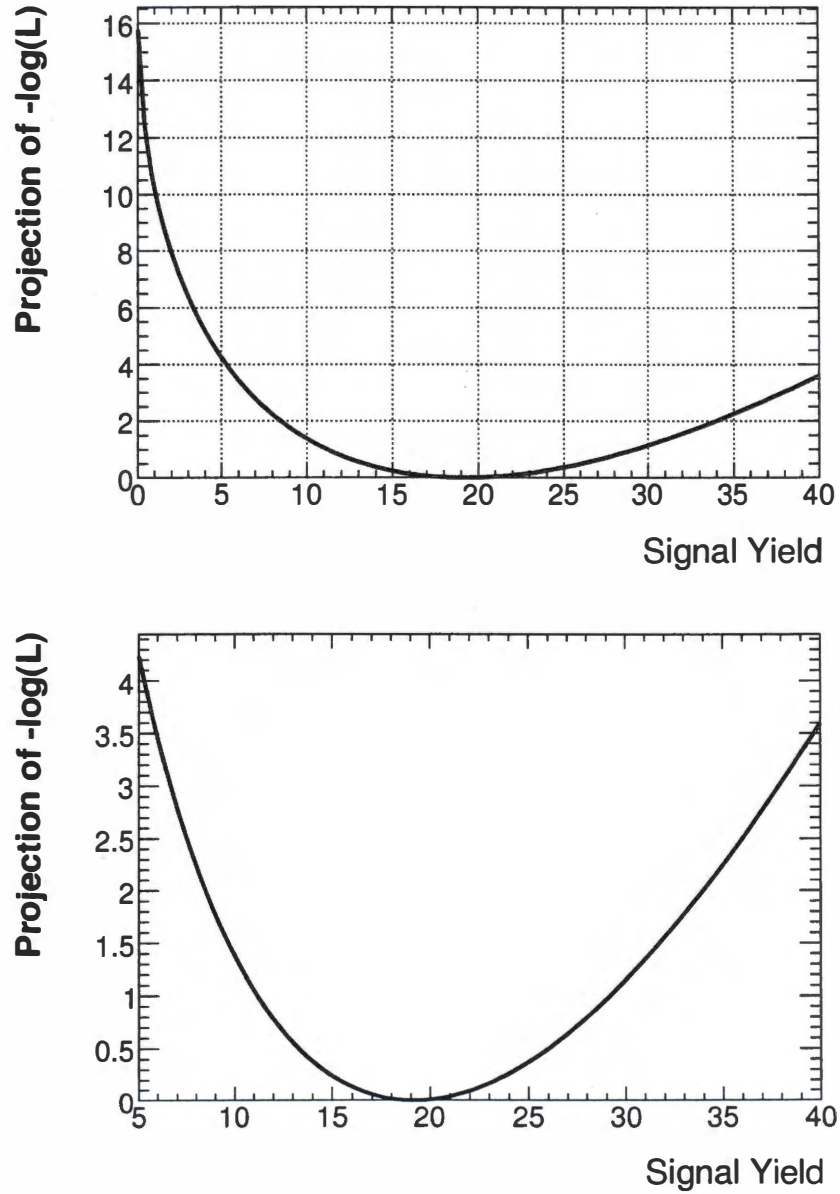


Figure 8.13: The $-\ln L/L_{max}$ of the yield fit PDF on data as a function of the number of signal events shows the statistical uncertainty of the fit at $N_{sig} = 0.0$ events. This corresponds to a 5.6 standard deviation statistical significance for the signal yield measurement.

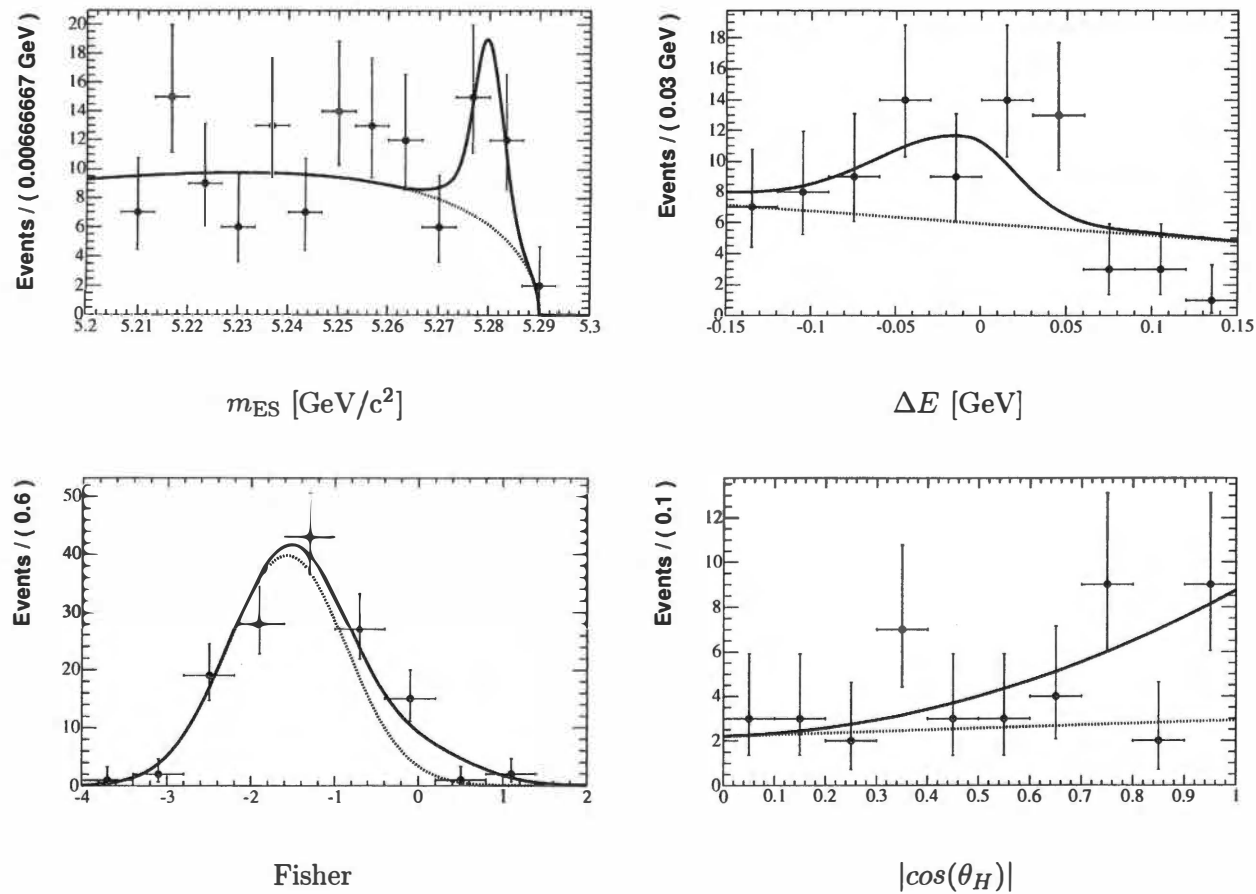


Figure 8.14: Likelihood-enhanced plots of the four event variables showing signal above the background level (dashed)

8.3 CP Analysis

8.3.1 Introduction

The time-dependent CP asymmetry between B^0 and $\bar{B}^0$ measures the CP -violation in the interference between mixing and decay. From the branching fraction analysis, the signal yield is 19 events. This is the maximum number of events that will be usable for the final CP fit. Since this is a very rare decay, the CP result for this channel alone will have a very low significance. Therefore, we choose to combine the $B^0 \rightarrow \phi K_S^0$, $K_S^0 \rightarrow \pi^0 \pi^0$ channel with the $K_S^0 \rightarrow \pi^+ \pi^-$ and ϕK_L^0 channels for the CP fit. This in turn enhances the overall significance of the CP measurement for $B^0 \rightarrow \phi K^0$.

The likelihood fit for the CP asymmetry in each channel is the product of the yield likelihood, as prepared in the previous section, and a likelihood that takes the proper time dependencies into account. Only part of the preparation of the latter and the support in its application was subject of this work. For completeness, we describe the full CP analysis in the following, though for details we refer to *BABAR* publications Ref. [18].

8.3.2 Strategy

In *BABAR*, neutral B mesons are produced in quantum-entangled pairs from the decay of the $\Upsilon(4S)$ resonance. As a consequence, the flavor of the two partners is only known at the moment one of the two decays. The overall time-dependent amplitude describing the propagation and decay of the system depends only on the difference in time between the two B decays. For the measurement of CP violation in the interference between mixing and decay, one B , B_{rec} , is fully reconstructed into the CP eigenstate ϕK^0 , while the other B , B_{tag} , is partially reconstructed to provide the timer start and the flavor tag at the time of its decay.

Figure 8.15 shows the schematic of a time-dependent CP measurement in *BABAR*. The z -axis indicates the beam direction. Electrons with an energy of 9 GeV move in the positive z direction, and positrons with 3.1 GeV of energy move in the negative z direction so that the direction of the resulting boost of $\beta\gamma = 0.56$ is in positive z . $\beta = v/c$ is the velocity v of the $\Upsilon(4S)$ in the laboratory frame divided by the speed-of-light, and γ is the relativistic dilation factor ($\gamma = 1/\sqrt{1 - \beta^2}$). Due to the low transverse momentum ($p_T < 300 \text{ MeV}/c$) of the B mesons in the $\Upsilon(4S)$ center-of-mass system (CMS), the B mesons have mostly longitudinal momentum along z and the transverse direction can be neglected (for demonstration purpose the figure exaggerates the transverse dimension). Due to the boost, the typical separation of decay vertices in z , $\Delta z = 30 \text{ } \mu\text{m}$, in the CMS, is stretched out to an average of $\Delta z = 250 \text{ } \mu\text{m}$. The proper decay time difference Δt

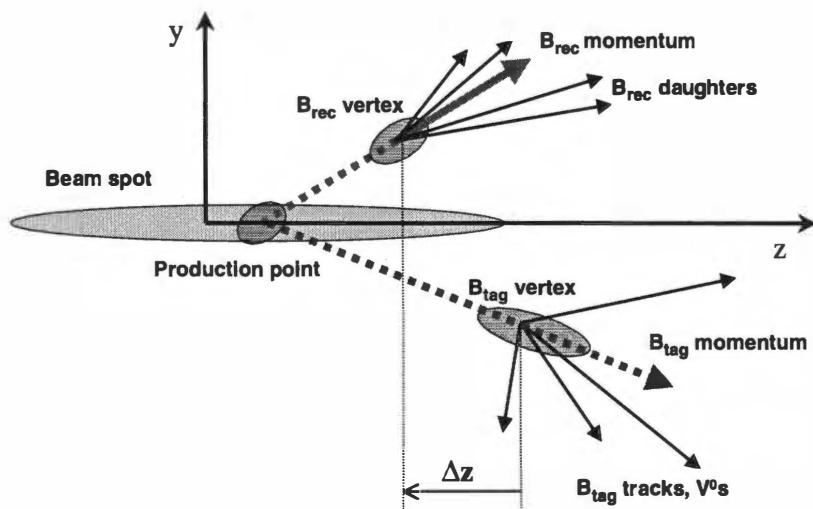


Figure 8.15: Schematic of the measurement of the time-dependent CP asymmetry in neutral B decays with the *BABAR* experiment

between the B_{rec} and B_{tag} decay vertices ($\Delta t = t_{\text{rec}} - t_{\text{tag}}$) in the event is calculated from the vertex separation Δz :

$$\Delta t = \Delta z / (\beta \gamma c). \quad (8.16)$$

As a consequence, the time-dependent measurement is independent from the primary production vertex, or beam spot, which has typically a much larger uncertainty than the vertices. The vertex reconstruction benefits from the increased decay lengths. The production point in Figure 8.15 is the result of a combined fit to both B vertices and the beam spot.

The CP asymmetry is defined as:

$$A(\Delta t) = \frac{N_{B^0}^{\text{tag}}(\Delta t) - N_{\bar{B}^0}^{\text{tag}}(\Delta t)}{N_{B^0}^{\text{tag}}(\Delta t) + N_{\bar{B}^0}^{\text{tag}}(\Delta t)}, \quad (8.17)$$

where $N_{B^0}^{\text{tag}}(\Delta t)$ is the number of events with a B tagged as B^0 at the time difference Δt , $N_{\bar{B}^0}^{\text{tag}}(\Delta t)$ is the corresponding number of events with $\bar{B}^0$ tags.

The decay time distributions $N_{B^0/\bar{B}^0}^{\text{tag}}(\Delta t)$ for the $B^0 \rightarrow \phi K_S^0$ decay signals are expected to have the following functional form:

$$f_{\pm}(\Delta t) = \frac{e^{-|\Delta t|/\tau}}{4\tau} [1 \pm S_f \sin(\Delta m_d \Delta t) \mp C_f \cos(\Delta m_d \Delta t)]. \quad (8.18)$$

The upper sign denotes a decay accompanied by a B^0 tag, the lower one by a $\bar{B}^0$ tag. The τ is the B^0 lifetime and Δm_d is the mixing frequency. The S_f and C_f are the parameters measuring CP violation in the interference between mixing and decay and direct CP violation, respectively. Assuming that the decay is dominated by Standard Model decay amplitudes, we expect $S_f \approx -\sin 2\beta$ and $C_f = 0$.

8.3.3 Measurement of Δt

In determining the location of the reconstructed B meson (B_{rec}) vertex, all of the charged tracks that make up B_{rec} are used. In this analysis, these are the two charged kaon tracks from the ϕ decay with some information from the K_S^0 direction. The tracks are fit to a common B vertex while varying track geometry parameters and momenta within their errors. The RMS resolution of the z position of the B_{rec} vertex is approximately $80 \mu\text{m}$ for 99% of the B_{rec} candidates. The tagging B meson vertex is formed by all charged tracks in the event that are not used in reconstructing B_{rec} . Electron and positron candidates that appear to originate from $\gamma \rightarrow e^+e^-$ conversions are rejected. The B_{tag} tracks are also subjected to a vertex fit. Tracks with large χ^2 contributions (> 6) are removed

one at a time, and the fit is repeated until all remaining tracks pass this requirement. This reduces the bias in the vertex position due to longer living D mesons. Finally, B_{rec} and B_{tag} vertices are fit together with the beam spot, within their resolution (see Figure 8.15), to extract Δz . The Δz RMS resolution of $190\,\mu\text{m}$ is limited by the B_{tag} vertex position resolution, as determined from Monte-Carlo simulations. The relative uncertainty introduced by misalignment of the vertex detector (SVT) is only about 0.2%. The boost factor in Eq. 8.16 is calculated from the beam energies, which are measured every 5 seconds, while keeping the relative uncertainty below 0.1%.

The measured Δt distributions for signal and background, $F_{\pm}(\Delta t)$, is described by a convolution of the real time evolution, $f_{\pm}(\Delta t)$, and the detector resolution function, $\mathcal{R}$, as:

$$F_{\pm}(\Delta t) = \int_{-\infty}^{\infty} d(\Delta t') \mathcal{R}(\Delta t - \Delta t') f_{\pm}(\Delta t'), \quad (8.19)$$

with the same sign convention as in Eq. 8.18. The $\mathcal{R}$ is a phenomenological function written in terms of the residual $\delta_t = \Delta t - \Delta t_{\text{true}}$, where Δt_{true} can be the true time difference from Monte-Carlo simulation. Nevertheless, the resolution function is dominated by the precision of the tagging vertex, so it is possible to extract its parameters from a fit to a large, fully reconstructed sample of B mesons with known B_{rec} flavor (see below for B_{flav} sample).

For *BABAR* CP measurements, the signal resolution function is typically represented by a sum of three Gaussian distributions: the core, tail, and outlier components. For the core and tail Gaussians, we use the measurement errors $\sigma(\Delta t)$ derived for each event but modified with a scale factor close to 1 that accommodates an overall under- or overestimate of the errors for all events. The presence of D mesons in the B_{tag} decays biases the Δt distribution in a small but systematic way. The D meson lifetime is long enough to displace somewhat the true B_{tag} vertex, because in the reconstruction we require that all decay products originate from one common vertex point. We, therefore, allow for non-zero means in the two central Gaussians. Furthermore, D mesons flying perpendicular to the z direction result in a somewhat better resolution than D mesons flying along the z direction in the laboratory. We take this correlation into account by scaling the mean offsets with the event-by-event error $\sigma(\Delta t)$. The third Gaussian accounts for fewer than 1% of events with incorrectly reconstructed vertices.

The background is dominantly combinatorial (continuum background), as established in the branching fraction analysis. It is expected to exhibit no time-dependence (prompt

decays) for (u, d, s) continuum events (a constant). The time dependence for $(\bar{c}c)$ continuum events with D meson production is $\exp(-|\Delta t|/\tau_c)$, where τ_c is the average charmed meson lifetime. The time resolution function for the background consists only of the core and the outlier Gaussian. The prompt background parameters in our charmless decay modes are floated on data.

8.3.4 B Flavor Tagging

The purpose of the “flavor tagging” algorithm is to determine the flavor of B_{tag} with the highest possible efficiency ϵ_{tag} and lowest possible probability w of assigning a wrong flavor to B_{tag} . This flavor tagging is achieved with the analysis of the decay products. The majority of the neutral B mesons decay into a final state that is flavor-specific, i.e., only accessible from either a B^0 or a $\bar{B}^0$, but not from both. An example is the semi-leptonic decay $B^0 \rightarrow D^{*-}l^+\nu_l$ or $\bar{B}^0 \rightarrow (D^{*+}l^-\bar{\nu}_l)$ with a branching fraction of about 5% [82]. The charge of the lepton unambiguously identifies the flavor of the neutral B . The subsequent decays of $D^{*+} \rightarrow D^0\pi^+$ and $D^0 \rightarrow K^-X^+$ result in a fast kaon in the final state whose charge also uniquely defines the B flavor after it is identified with information from the Cherenkov detector. Another example is the self-tagging decay $B^0 \rightarrow D^{*-}\pi^+$. This decay produces an energetic π^+ and a low-momentum π^- in the final state, and the B_{tag} flavor can be inferred from the charges of both.

The *BABAR* tagging algorithm is based on a neural network and assigns each event to one of mutually exclusive tag categories such as Lepton, Kaon, Kaon-Pion, slow-Pion, or Other. The name of each category indicates the dominant physics processes contributing to the flavor identification. Each category has different tagging efficiency and mis-tag fraction and different Δt resolution that can be established from the large statistics B flavor sample (see below).

Differences in the mis-tag fractions between B and $\bar{B}$ can arise from the different response of the detector to positively charged pions and kaons and negatively charged pions and kaons (because they have different total and charge-exchange cross sections). To account for possible differences, separate mis-tag probabilities, w for B and $\bar{w}$ for $\bar{B}$, are introduced. We define:

$$\langle w \rangle = \frac{1}{2}(w + \bar{w}) \quad \text{average mis-tag probability,} \quad (8.20)$$

$$\Delta w = (w - \bar{w}) \quad \text{mis-tag probability difference.} \quad (8.21)$$

With these definitions, the Δt distribution (Eq. 8.18), including flavor tagging, becomes:

$$f_{\pm} = \frac{f^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} [1 \mp \Delta w \pm (1 - 2\langle w \rangle)[S_f \sin(\Delta m_d \Delta t) - C_f \cos(\Delta m_d \Delta t)]]. \quad (8.22)$$

Notice that when the flavor tag is unresolved ($w = 0.5$), CP violation cannot be measured. The Figure of Merit for the performance of the tagging algorithm is the analyzing power Q , defined as:

$$Q = \epsilon_{\text{tag}}(1 - 2\langle w \rangle)^2. \quad (8.23)$$

This is related to the statistical uncertainty σ in the coefficients S_f and C_f :

$$\sigma \propto \frac{1}{\sqrt{Q}}. \quad (8.24)$$

8.3.5 B Flavor Sample

Wrong tag probabilities in the different tagging categories and the parameters of the time difference resolution function are measured with a sample of about 46,000 fully reconstructed B^0 decays, B_{flav} , in hadronic flavor eigenstate modes involving $D^{(*)}\pi/\rho/a_1$ and $J/\psi K^*(K^* \rightarrow K^+\pi^-)$. The decay products of $B_{\text{flav}} = B_{\text{rec}}$ indicate its flavor, the other B in the event, B_{tag} , is partially reconstructed and tagged by its decay products as before. The reconstruction and selection, with respect to vertex and flavor tag, is identical to the CP sample. For the neutral B system produced from $\Upsilon(4S)$, the probability for obtaining a mixed $B^0 B^0$ or $\bar{B}^0 \bar{B}^0$, or unmixed $B^0 \bar{B}^0$ final state is a function of the oscillation frequency Δm_d and the proper time difference Δt between the two B decays. The probability density function for the mixed (-) and unmixed (+) events, with the tag being a B^0 or $\bar{B}^0$, is:

$$\begin{aligned} g_{\pm, \text{tag}=B^0} &= \frac{e^{-|\Delta t|/\tau}}{4\tau} [1 + \Delta w \pm (1 - 2\langle w \rangle) \cos(\Delta m_d \Delta t)], \\ g_{\pm, \text{tag}=\bar{B}^0} &= \frac{e^{-|\Delta t|/\tau}}{4\tau} [1 - \Delta w \pm (1 - 2\langle w \rangle) \cos(\Delta m_d \Delta t)], \end{aligned} \quad (8.25)$$

with Δw and $\langle w \rangle$ from Eq. 8.21. It is convoluted with the same time-difference resolution function as for the CP measurement Eq. 8.19. The conformity between flavor sample and CP sample has been established by the BABAR collaboration from data and MC samples of fully reconstructed $B\bar{B}$ events.

The purity of the flavor sample is high ($> 90\%$), and residual background is mostly combinatoric in nature. The tagging efficiency summed over all tagging categories is $\epsilon_{\text{tag}} \approx 74\%$, and the corresponding analyzing power is $Q = 30.6 \pm 0.4\%$ [82].

8.3.6 Maximum Likelihood Fit

We simultaneously fit the three CP channels ϕK_S^0 , $K_S^0 \rightarrow \pi^+\pi^-$ (4300 candidate events), ϕK_L (8238 candidate events), and ϕK_S^0 , $K_S^0 \rightarrow \pi^0\pi^0$ (1129 candidate events) together with the B flavor sample ($\approx 46,000$ events) to extract the CP parameters S_f and C_f , where S_f flips sign in ϕK_L . The probability density function $\mathcal{P}_{i,c}^j$ for each event j is a product of the PDFs for the separate observables. For each event hypothesis i (signal or background) and tagging category c , we define $\mathcal{P}_{i,c}^j = \mathcal{P}_i(m_{ES}) \cdot \mathcal{P}_i(\Delta E) \cdot \mathcal{P}_i(\mathcal{F}) \cdot \mathcal{P}_i(\cos \theta_H) \cdot \mathcal{P}_i(\Delta t; \sigma_{\Delta t}, c)$, where, for the ϕK_L^0 mode, $\mathcal{P}_i(m_{ES}) = 1$ and, for the flavor sample, $\mathcal{P}_i(\mathcal{F}) \cdot \mathcal{P}_i(\cos \theta_H) \cdot \mathcal{P}_i(m_{KK}) = 1$. The $\sigma_{\Delta t}$ is the error on Δt for a given event. The likelihood function for each decay chain is then

$$L = \prod_c \exp \left(- \sum_i N_{i,c} \right) \prod_j^{N_c} \left[\sum_i N_{i,c} \mathcal{P}_{i,c}^j \right], \quad (8.26)$$

where $N_{i,c}$ is the yield of events of hypothesis i found by the fitter in category c (flavor or charge), and N_c is the number of category c events in the sample. Without the CP part, we extract the number of signal candidates from each sample: 130 in the combined ϕK_S^0 channels and 98 in ϕK_L .

8.3.7 Studies of the Likelihood Fit

In the same manner as for the branching fraction fit, we establish the reliability and potential biases of the likelihood fit with samples from MC simulated events and events generated from the PDFs (toy MC). This is a computing intense task and we make full use of the computer cluster.

The samples use signal events with the expected numbers in the different decay channels randomly picked from reconstructed SP6 MC. With the CP parameters unknown, several different samples for different S_f and C_f combinations are prepared. We also select events from samples of potential background from mis-reconstructed $B^0\bar{B}^0$ in the expected amount. Continuum background is generated for each sample independently from the PDFs and according to the expected yields. In all tests, the simulated values are well reproduced.

For the final fit result, the procedure is repeated. We generate 1000 samples and obtain:

	mean error	pull mean	pull width
S_f	0.27	0.03 ± 0.03	1.04 ± 0.03
C_f	0.23	-0.02 ± 0.03	1.03 ± 0.03

We observe no significant bias, and the errors from the simulation agree with the errors from the fit.

8.3.8 Systematic Error

The uncertainty of the fit is dominated by the statistical error. Parameters for tagging and time difference resolution are extracted simultaneously from a large flavor data set and only the uncertainty due to the choice of the models for the PDFs (for signal and background) need to be taken into account. Alternative parameterizations (e.g., 2 instead of 3 Gaussians) have been tested, and the variation of the CP parameters is considered as the uncertainty. The largest uncertainty in the parameterization of the signal yield originates from the fixed choice of the endpoint in the variable m_{ES} . We vary the point within one standard deviation and refit. The shift in the values of S_f and C_f is accounted for as systematic error. Another uncertainty arises from possible neutral B decay background, which can distort the CP asymmetry. Though expected to be small, we vary its content within the uncertainty of the branching fractions and assume maximal CP violation if not known. Finally, we measure the tagging parameters from a sample of fully reconstructed B decays, and use charges of decay products to determine the flavor of the two B mesons. When the final state originates from a $b \rightarrow c\bar{u}d$ transition, we assume that this is the only amplitude. However, in some fraction of all events, it can also originate from a doubly-CKM-suppressed $\bar{b} \rightarrow \bar{u}c\bar{d}$ transition. Neglecting the suppressed amplitude introduces a bias in the Δt distribution and, consequently, in the asymmetry coefficients. We evaluate upper limits on these deviations using the formalism described in Ref. [83]. The total error is obtained by summing the individual errors in quadrature.

8.3.9 Result

From the maximum likelihood fit we measure:

$$S_{\phi K^0} = 0.48 \pm 0.27(\text{stat}) \pm 0.06(\text{syst}), \quad (8.27)$$

$$C_{\phi K^0} = 0.00 \pm 0.23(\text{stat}) \pm 0.05(\text{syst}). \quad (8.28)$$

The correlation among the CP parameters is less than 6%. The likelihood from the fit to data compares well with the likelihood from the mock fits in the previous section (which is an indication of the goodness of the fit). We compare those values to the

measurement in the neutral B decays to charmonium final states [23] that are dominated by tree-amplitudes and hence are close to the Standard Model value:

$$S_{c\bar{c}} = 0.71 \pm 0.03(\text{stat}) \pm 0.02(\text{syst}), \quad (8.29)$$

$$C_{c\bar{c}} = 0.07 \pm 0.03(\text{stat}) \pm 0.02(\text{syst}). \quad (8.30)$$

The CP parameter values from the gluonic penguin modes ϕK^0 agree within one standard deviation with the Standard Model values from the combined $c\bar{c}$ modes.

Figure 8.16 shows the Δt distributions of the B^0 -tagged and $\bar{B}^0$ -tagged subsets together with the raw asymmetry, for ϕK_S^0 and ϕK_L^0 events separately, with the result of the combined time-dependent CP -asymmetry fit superimposed.

In conclusion, we do not observe a significant deviation of the CP asymmetry from Standard Model expectation (which would have been an indication of new particles and forces beyond the Standard Model). If the central values are accurate, in the ϕK^0 channels alone there will be not enough statistics during the lifetime of *BABAR* (end of 2008) to arrive at a significant deviation.

8.4 Conclusion

We observe the decay $B^0 \rightarrow \phi K_S^0$, $\phi \rightarrow K^+ K^-$, $K_S^0 \rightarrow \pi^0 \pi^0 \rightarrow 4\gamma$ with a statistical significance of 5.6σ and measure the branching fraction of $B^0 \rightarrow \phi K^0$ to be

$$BF(B^0 \rightarrow \phi K^0) = (6.6_{-2.0}^{+2.4}(\text{stat}) \pm 0.42(\text{syst})) \times 10^{-6}. \quad (8.31)$$

We also measure the CP violating asymmetry in the combined $B^0 \rightarrow \phi K^0$ final states. The asymmetry is

$$S_{c\bar{c}} = 0.48 \pm 0.27(\text{stat}) \pm 0.06(\text{syst}), \quad (8.32)$$

$$C_{c\bar{c}} = 0.00 \pm 0.23(\text{stat}) \pm 0.05(\text{syst}). \quad (8.33)$$

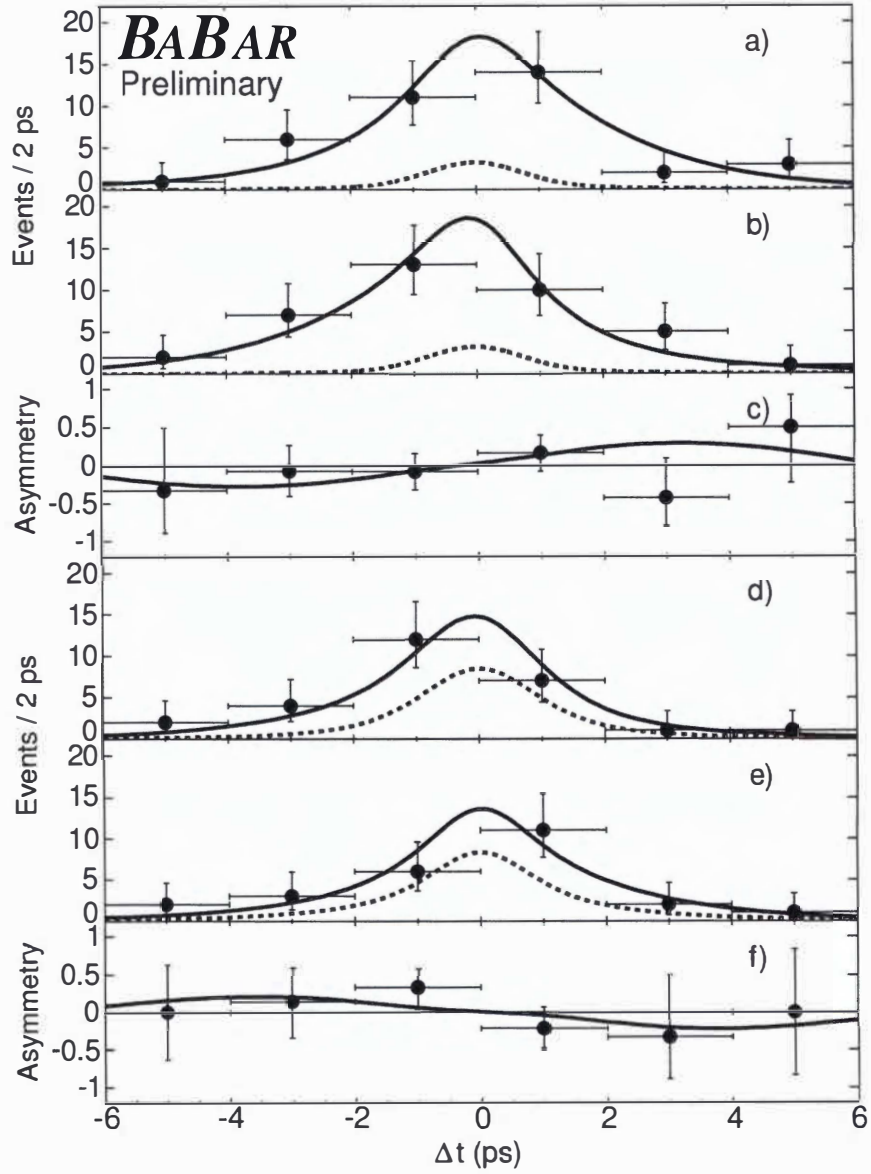


Figure 8.16: Plots (a) and (b) show the Δt distributions of B^0 and $\bar{B}^0$ -tagged ϕK_s^0 candidates. The solid lines refer to the fit for all events; the dashed lines correspond to the background. Plot (c) shows the raw asymmetry. Plots (d), (e), and (f) are the corresponding plots for ϕK_L^0 candidates. For each final state, a requirement is applied on the event likelihood to suppress background.

Bibliography

Bibliography

- [1] J. Schwindling, B. Mansoulie, MLPfit: a tool for designing and using multi-layer perceptrons (January 2000).
URL <http://schwind.home.cern.ch/schwind/MLPfit/doc/MLPfit.pdf>
- [2] D. Griffiths, Introduction to Elementary Particles, Wiley, 1987.
- [3] S. F. Novaes, Standard Model: An Introduction (2000).
URL <http://arxiv.org/abs/hep-ph/0001283>
- [4] http://nobelprize.org/nobel_prizes/physics/laureates/1999/.
- [5] W. M. Yao, et al., Review of Particle Physics, Journal of Physics G 33 (2006) 1+.
URL <http://pdg.lbl.gov/>
- [6] S. Martin, A Supersymmetry Primer (2006).
URL <http://arxiv.org/abs/hep-ph/9709356>
- [7] M. Turner, Cosmology solved? Quite possibly!, Publ. Astron. Soc. Pac. 111 (1999) 264–273.
- [8] E. Noether, Invariante Variationsprobleme, Math.-Phys. Kl. (1918) 186–207.
- [9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, R. P. Hudson, Experimental Test of Parity Conservation in Beta Decay, Phys. Rev. 105 (4) (1957) 1413–1415.
- [10] B. Kayser, Neutrino Mass, Mixing, and Flavor Change, in [23].
- [11] J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay, Evidence for the 2π Decay of the K_2^0 Meson, Phys. Rev. Lett. 13 (4) (1964) 138–140.
- [12] L. Wolfenstein, Violation of CP Invariance and the Possibility of Very Weak Interactions, Phys. Rev. Lett. 13 (18) (1964) 562–564.

- [13] G. Lüders, On the equivalence of invariance under time reversal and under particle-antiparticle conjugation for relativistic field theories, *Mat.-fys. Medd.* 28 (5) (1954) 1–17.
- [14] W. Pauli, L. Rosenfeld, V. Weisskopf, Niels Bohr and the Development of Physics, McGraw-Hill, New York, 1955.
- [15] J. S. Bell, CPT Violation and the Standard Model, Ph.D. thesis, Birmingham University (1954).
- [16] R. Jost, Eine Bemerkung zum CTP Theorem, *Helvetica Physica Acta* 30 (1957) 409–416.
- [17] A. D. Sakharov, *pisma Zh. Eksp. Teor. Fiz* (1967).
- [18] B. Aubert, et al., Study of time-dependent CP-violating asymmetries and flavor oscillations in neutral B decays at the $\psi(4S)$, *Phys. Rev. D* 66 (3) (2002) 032003.
- [19] S. Barr, G. Segrè, H. A. Weldon, Magnitude of the cosmological baryon asymmetry, *Phys. Rev. D* 20 (10) (1979) 2494–2498.
- [20] M. Kobayashi, T. Maskawa, CP-Violation in the Renormalizable Theory of Weak Interaction, *Prog. Theor. Phys.* 49 (2) (1973) 652–657.
- [21] N. Cabibbo, Unitary Symmetry and Leptonic Decays, *Phys. Rev. Lett.* 10 (12) (1963) 531–533.
- [22] F. Gilman, et al., The CKM Quark Mixing Matrix, *Physics Letters B* 592 (2004) 793.
- [23] S. Eidelman, et al., Review of Particle Physics, *Physics Letters B* 592 (2004) 1+. URL <http://pdg.lbl.gov/>
- [24] C. Jarlskog, Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation, *Phys. Rev. Lett.* 55 (10) (1985) 1039–1042.
- [25] C. O. Dib, I. Dunietz, F. J. Gilman, Y. Nir, Standard-model predictions for CP violation in B^0 -meson decay, *Phys. Rev. D* 41 (5) (1990) 1522–1530.

- [26] J. Ellis, M. K. Gaillard, D. V. Nanopoulos, S. Rudaz, The phenomenology of the next left-handed quarks, *Nuclear Physics B* 131 (1977) 285–307.
- [27] K. Lingel, T. Skwarnicki, J. G. Smith, Penguin decays of B mesons, *Ann. Rev. Nucl. Part. Sci.* 48 (1998) 253–306.
- [28] e. Harrison, P. F., e. Quinn, Helen R., The BaBar physics book: Physics at an asymmetric B factory, papers from Workshop on Physics at an Asymmetric B Factory (BaBar Collaboration Meeting), Rome, Italy, 11-14 (Nov 1996), Princeton, NJ, 17-20 (Mar 1997), Orsay, France, 16-19 (Jun 1997) and Pasadena, CA, 22-24 (Sep 1997).
- [29] B. Aubert, et al., Measurement of B^0 -anti- B^0 Flavor Oscillations in Hadronic B^0 Decays, *Phys. Rev. Lett.* 88 (22) (2002) 221802.
- [30] B. Aubert, et al., Direct CP Violating Asymmetry in $B^0 \rightarrow K^+\pi^-$ Decays, *Physical Review Letters* 93 (13) (2004) 131801.
URL <http://link.aps.org/abstract/PRL/v93/e131801>
- [31] R. A. Briere, et al., Observation of $B \rightarrow \phi K$ and $B \rightarrow \phi K^*$, *Phys. Rev. Lett.* 86 (17) (2001) 3718–3721.
- [32] B. Aubert, D. Boutigny, J.-M. Gaillard, A. Hicheur, Y. Karyotakis, J. P. Lees, P. Robbe, V. Tisserand, A. Palano, G. P. Chen, J. C. Chen, N. D. Qi, G. Rong, P. Wang, Y. S. Zhu, G. Eigen, P. L. Reinertsen, B. Stugu, B. Abbott, G. S. Abrams, A. W. Borgland, A. B. Breon, D. N. Brown, J. Button-Shafer, R. N. Cahn, A. R. Clark, Q. Fan, Measurement of the Decays $B \rightarrow \phi K$ and $B \rightarrow \phi K^*$, *Phys. Rev. Lett.* 87 (15) (2001) 151801.
- [33] P. J. Oddone, Linear Collider B-Factory Conceptual Design, in: D. Stork (Ed.), *Proceedings of the UCLA Workshop*, World Scientific, 1987, p. 243.
- [34] A. Abashian, et al., The Belle Detector, *Nucl. Instr. Meth. A* 479 (1) (2002) 117–232.
- [35] B. Aubert, et al., The BaBar Detector, *Nucl. Instr. Meth. A* 479 (1) (2002) 1–116.
- [36] BaBar Collaboration, The BaBar Detector, Tech. Rep. SLAC-PUB-8569, SLAC (April 2001).
- [37] H. Bichsel, D. Groom, S. Klein, Electronic energy loss by heavy particles, in [5].

- [38] L. Lyons, Statistics for nuclear and particle physicists, Cambridge University Press, 1986, Ch. 4,6, pp. 85–98,161–168.
- [39] S. Agostinelli, et al., GEANT4 – a simulation toolkit, Nucl. Instr. Meth. A 506 (2003) 250–303.
- [40] Intel Inc., Preboot Execution Environment (PXE) Specification (September 1999).
URL <http://www.pix.net/software/pxeboot/archive/pxespec.pdf>
- [41] W. Stevens, TCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley, 1994.
- [42] P. Chen, E. Lee, G. Gibson, R. Katz, D. Patterson, RAID: High-Performance, Reliable Secondary Storage, ACM Computing Surveys 26 (2) (1994) 145–185.
- [43] EIA-TSB-67, as supplemented by EIA ANSI/TIA/EIA-568-B. 2-1.
- [44] C. Heine, Swapping via NFS for Linux (March 2005).
URL <http://www.nfs-swap.dot-heine.de/>
- [45] Objectivity, Inc., Menlo Park, CA, Objectivity Database System Overview (1990).
- [46] M. Steinke, et al., How to build an event store – the new Kanga event store for BaBar, in: Proceedings of CHEP 2004, Interlaken, Switzerland, 2004.
- [47] R. Brun, F. Rademakers, ROOT – an object oriented data analysis framework, Nucl. Inst. Meth. A 389 (1997) 81–86.
- [48] R. Brun, et al., PAW – physics analysis workstation, CERN program library Q121, CERN, Geneva, Switzerland (1989).
- [49] M. Goto, cINT C/C++ interpreter.
URL <http://root.cern.ch/root/Cint.html>
- [50] W. Verkerke, D. Kirkby, The RooFit toolkit for data modeling, physics/0306116.
- [51] <http://www.slac.stanford.edu/BFR00T/www/Computing/Environment/Tools/RequiredSoft/>.
- [52] CERN Program Library.
URL <http://cernlib.web.cern.ch/>
- [53] Free Software Foundation, Inc., The GNU Project.
URL <http://www.gnu.org/>

- [54] L. Wall, T. Christiansen, R. Schwartz, Programming Perl, Third Edition, O'Reilly, 2000.
- [55] J. K. Ousterhout, Tcl and the Tk Toolkit, Addison-Wesley, 1994.
- [56] Wireless Application Protocol WAP 2.0 Technical White Paper, Wireless Application Protocol Forum, January 2002.
- [57] <http://babar2.phys.utk.edu/~gragghia/scripts.cgi?id=dwarf>.
- [58] P. H. Carns, et al., PVFS: A Parallel File System for Linux Clusters, in: Proc. of the Extreme Linux Track: 4th Annual Linux Showcase and Conf., 2000.
- [59] A. Shishani, A. Sim, J. Gu, Storage Resource Managers: Middleware Components for Grid Storage, in: Nineteenth IEEE Symposium on Mass Storage Systems, 2002.
- [60] P. Fuhrmann, dcache, the commodity cache, in: Twelfth NASA Goddard and Twenty First IEEE Conference on Mass Storage Systems and Technologies, Washington, DC, 2004.
- [61] D. Brown, E. Charles, D. Roberts, The BaBar Track Fitting Algorithm, in: Proceedings of CHEP 2000, Padova, Italy, 2000.
- [62] A. Drescher, B. Gräwe, B. Hahn, B. Ingelbach, U. Matthiesen, H. Scheck, J. Spengler, D. Wegener, The Argus electron-photon calorimeter III, Nucl. Instr. Methods A237 (3) (1985) 464–474.
- [63] J. Brose, S. R. Freund, K. Schubert, T. U. Dresden, A Reconstruction Method for KS to $\pi^0 \pi^0$ Decays and its Application to Bz to Jpsi KS ($\pi^0 \pi^0$) Events, BaBar Analysis Document #221 (October 2001).
- [64] R. Aleksan, I. Dunietz, B. Kayser, F. Le Diberder, CP violation using non-CP eigenstate decays of neutral B mesons, Nuclear Physics B 361 (1991) 141+.
- [65] G. C. Fox, S. Wolfram, Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes, Phys. Rev. Lett. 41 (23) (1978) 1581–1585.
- [66] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (1936) 178–188.
- [67] F. Rosenblatt, The Perceptron: A Probabalistic Model for Information Storage and Organization in the Brain, Psychological Review 65 (6) (1958) 386–408.

- [68] M. L. Minsky, S. Papert, Perceptrons: an introduction to computational geometry, MIT Press, Cambridge, 1969.
- [69] R. Barlow, Introduction to Statistical Issues in Particle Physics (2003).
URL <http://arxiv.org/abs/physics/0311105>
- [70] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks as universal approximators, Neural Netw. 2 (5) (1989) 359–366.
- [71] H. Robbins, S. Monro, A Stochastic Approximation Method, Annals of Math. Stat. 22 (1951) 400.
- [72] R. A. Fisher, F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, Oliver and Boyd, London, 1938.
- [73] M. Krishnamurthy, Study of $B \rightarrow \phi K_L$, presented at ICHEP 2004 conference, published in Phys. Rev. D (2005) 0911012 (2003).
- [74] Bottom Mesons: B0: Decay Modes, in [5].
- [75] J. Klein, A. Roodman, Blind analysis in nuclear and particle physics, Annual Review of Nuclear and Particle Science 55 (1) (2005) 141–163.
- [76] H. Albrecht, Argus Collaboration, Z. Phys. C48 (1990) 543.
- [77] F. James, M. Roos, Minuit – a system for function minimization and analysis of the parameter errors and correlations, Computer Physics Communications.
- [78] B. Aubert, et al., A study of time-dependent CP-violating asymmetries in $B^0 \rightarrow J/\psi K_S^0$ and $B^0 \rightarrow \psi(2S) K_S^0$ decays (2000).
URL <http://arxiv.org/abs/hep-ex/0008048>
- [79] M. Tiller, M. Naisbit, A. Roodman, A Study of π^0 Efficiency, BaBar Analysis Document #870 (2004).
- [80] Studies on π^0 Reconstruction, BaBar Analysis Document #20 (2000).
- [81] G. Raymond, et al., Luminosity Measurement for the Runs 1,2 and 3 data sample using Release 12 and SP5 Simulation, BaBar Analysis Document #1312 (2005).
- [82] M. Bruinsma, D. Lange, G. Sciolla, M. Simani, A B flavor tagging algorithm for CP violation measurements at the BaBar experiment, Tech. Rep. BAD1025, The BaBar Collaboration (2005).

- [83] O. Long, M. Baak, R. N. Cahn, D. Kirkby, Impact of tag-side interference on time-dependent CP asymmetry measurements using coherent $B^0\bar{B}^0$ pairs, Physical Review D (Particles and Fields) 68 (3) (2003) 034010.
URL <http://link.aps.org/abstract/PRD/v68/e034010>

Vita

Gerald C. Ragghianti Jr. was born in Memphis, Tennessee, the son of Gerald and Sharon Ragghianti. After graduating in 1998 from Craigmont High School optional program, he attended The University of Tennessee as a Chancellor's Scholar and graduated *cum laude* in 2002 with a Bachelor of Science degree in physics. As an undergraduate, he participated in research at the Oak Ridge National Laboratory in instrumentation for Condensed Matter and co-authored a paper in the journal Physical Review B. He received a Science Alliance Fellowship to begin graduate studies at The University of Tennessee in physics and joined the High Energy group in 2003 to develop a computing system for *BABAR* Monte-Carlo simulation. In 2004, he presented the *BABAR* group's work in *CP* violation at the Southeastern Section APS meeting in Oak Ridge, Tennessee. Following completion of his Masters Degree in Physics, Gerald plans to start work in creating a computing center at The University of Tennessee for the upcoming CMS experiment in Geneva, Switzerland.

