A Study of BaBar Detector Behaviors Using the $e^+e^- \to \varphi\gamma$ Control Sample

By

Nasim Boustani

B.A.Sc., Shiraz University of Iran, 2000

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In

The Faculty of Graduate Studies

(Physics)

THE UNIVERSITY OF BRITISH COLOMBIA

April 20, 2005

Abstract

This thesis uses the control sample $e^+e^- \to \varphi\gamma, \varphi \to K^+K^-$ to presents the study of extra energy deposited in the calorimeter in BaBar detector and a comparison between data and Monte Carlo to see how well the behavior is simulated in simulated data. It also presents the study of the kaon identification system efficiency currently used in BaBar experiment.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to give special thanks to my supervisor Dr. Christopher Hearty for all his help, guidance and patience. I owe him more than I can thank for for the completion of this study. I would also like to thank Dr. Janis McKenna and Dr. Thomas Mattison for patiently answering all my questions during this time. Thanks to Dr. Chris Waltham for kindly understanding my time constraints for submission deadline, and reviewing my study in quite a short period of time.

Thanks to all UBC BaBar group, especially David Asgeirsson for spending so much of his time proofreading this thesis, which is my first publication in English. He did a lot for me and I cannot thank him enough for that. Thanks to Dr. Douglas Thiessen for helping me familiarize with BaBar system and always saving me from Linux confusions.

This acknowledgement would not be nearly complete without thanking all my friends for their valuable help and kindness, especially Ali Reza Mohazzab. He did the first stage of proofreading of almost the entire document and guided me through all Microsoft Word ups and downs. Many thanks to my dear friends Shirin Hadizadeh and again David Asgeirsson for graciously accepting to go through all the waiting to submit this thesis on behalf of me.

There is no way to thank my family enough for their immense support, specially my mom and dad, who always believed in me, and have always guided me through hardship in my life.

Great thanks to my guru and mentor, Shahzad Pakzad, for being a channel to unimaginable greatness and beauty.

I sincerely thank everyone, everywhere whom their help and support has helped completion of this part of my life.

I thank god for everything.

A

Introduction

Mankind has always been challenged to understand nature. For as long as we can remember, we have always been looking for answers which in turn have led us to whole new sets of questions yet to be asked. Curiosity is the driving force of the human mind. With enough motivation and effort, you might find yourself spending your whole life trying to answer some of the questions you had in your mind as a child.

Let's suppose you are 10 years old and after having your breakfast you start cutting a piece of muffin in half and that piece in half and so on. After a while, you begin to ponder is there an end to this process? If you continuously cut pieces in half, how small of a piece will you end up with at the end? What are the fundamental building blocks of things around us?

Well, this question has long been pondered before you started cutting that muffin. Today, the theory that may provide some answers to these sorts of questions is called "The Standard Model of Particle Physics". According to this model, all matter in the universe is made up of twelve particles: six leptons and six quarks.

				Charge
Leptons	e (electron)	μ (muon)	τ (tau)	-1e
	υ_e (electron neutrino)	υ_μ (muon neutrino)	υ_τ (tau neutrino)	0
Quarks	u (up)	c (charm)	t (top)	+2/3e
	d (down)	s (strange)	b (bottom)	-1/3e

Table A-1: Quarks and leptons. All the visible matter in the universe is supposed to be made of a combination of these particles.

Each of these elementary particles has a counterpart, an antiparticle, which can be thought of as the building blocks of antimatter. Antiparticles have the same mass as their particle twin, but the opposite sign for some quantum numbers like charge. Along with the particles, there are four known forces in the world which allow these particles to interact with each other. These forces are: the Gravitational force (G), the

Electromagnetic force (EM), the Weak Nuclear force (W) and the Strong Nuclear force (S).

It was long thought that the forces of nature preserve symmetry between matter and antimatter. In general, a theory of physics is said to have a symmetry if its laws apply equally well even after some operation, such as reflection or parity reversal (P), transforms parts of the physical system.

An important example is the parity reversal operation (P) which turns an object into its mirror reflection and rotates it 180 degrees about the axis perpendicular to the mirror. (Figure A-1)

In 1957 Madame Wu and her collaborators showed that the weak force totally violates P symmetry[1]. Later on, it was also shown that symmetry under the charge conjugation operator, which changes the quantum numbers of each particle into those of its antiparticle, is also violated by the Weak force[1].

It seemed, however, that the forces would still be symmetric under the combination of C and P, which is actually the correct operation to convert a particle into its antiparticle. In other words it was thought that nature didn't have a way of distinguishing between left-handed matter and right-handed antimatter.

But, just by looking at the sky above you, you might suspect that there is something wrong with these hypotheses. For as far as astronomers can see into the universe, everything is made entirely of matter. Essentially no antimatter is evident. If the known laws of physics cannot distinguish between matter and antimatter, then what has caused this huge imbalance between the two in the universe around us?

In 1964 an experiment at Brookhaven National Laboratory proved that in weak interaction, CP too is not a conserved quantity [2]. These results severely contradicted the Standard Model at that time. In 1972 Kobayashi and Maskawa proposed that if there are three or more generations of quarks, then CP symmetry could be violated within the Standard Model [3]. The third generation was indeed found experimentally in the following years, but one problem still remained: The amount of CP asymmetry observed in nature could not account for even a billionth of what is necessary to have produced our current universe from a symmetric one right after the big bang[2]. So the search for mechanisms that violate CP-symmetry continued.

---

[1] The C operator should make an antiparticle out of a particle. This is not true as we can see by examining neutrinos (Figure A-1). Neutrinos are exclusively left-handed particles. That is, the spin of a neutrino is always in the opposite direction of its velocity. Antineutrinos on the other hand, are always right-handed particles. If we applied the C operator on a neutrino, we would get a left-handed antineutrino, which does not exist (this is one of the reasons that the C is not a good symmetry of nature). Also the P operator generates a right-handed neutrino out of a left-handed one, which is again impossible. To be able to get a real antineutrino, we have to combine C and P, which would make a right-handed antineutrino out of a left-handed neutrino.

[2] The amount of CP asymmetry needed is indeed very tiny itself. It would be sufficient that for each 10^9 antiparticles, there would be 10^9+1 particles some time after big bang. The 10^9 parts would cancel each other out, leaving behind a considerable amount of background energy, and the 1 remaining part of matter would make up our astonishing universe.

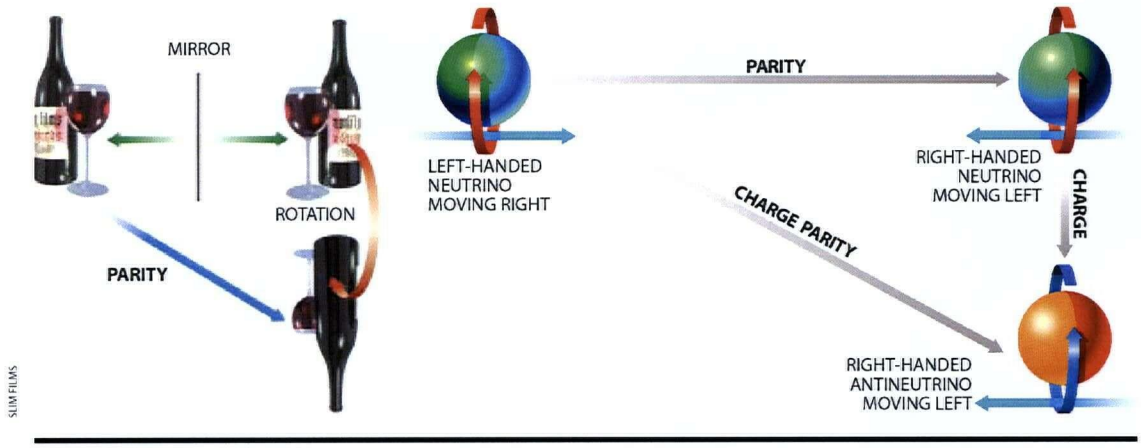

Figure A-1: Parity reversal and Charge conjugation operators.

One fruitful place to search for more CP-violation is in the B mesons. The standard model predicts the various decays of the B^0 and $\overline{B}^0$ to be highly asymmetric.

These predictions called for the construction of a so-called "B-Factory", a place where millions of B mesons are produced to be studied. The name BaBar comes from B B bar. It is a detector placed right around the interaction point of two high-energy beams of electrons and positrons, tuned to make as many B s as possible. The goal of the BaBar experiment is to provide an answer to this mind-boggling mystery: Why is the universe so lopsided? What caused this great triumph of matter over antimatter?

To be able to study what really is happening in the B meson system produced in the detector, one must know with a high degree of accuracy how the detector works. This study is an effort to shine some light on this matter. It will discuss how the detector responds in some specific situations and also how well this behavior is simulated in the simulation programs. I hope this study can provide some help for someone looking at the bigger picture of this puzzle.

1

The BaBar Detector

1.1 Introduction

The primary goal of the BaBar experiment is the detection and measurement of CP violation in the B - meson system. To facilitate this, we need to be able to distinguish between the two opposite B flavors. This is done by boosting the B mesons in the lab frame, causing a separation of their decay vertices that can be measured by existing silicon vertex detector technology. In order to achieve this, the energy of the electron beam and the positron beam are unequal. The fact that the decay products are boosted forward in the lab frame creates a need for an asymmetric detector. This chapter will begin by describing the asymmetric collider on which the BaBar detector is situated. The detector itself will then be described briefly. For more extensive discussion you can refer to [5], [6] and [7].

1.2 The Linac

Electrons produced by a heated filament are accelerated in a two-mile Linac to 9 GeV using electromagnetic fields in cavities along the accelerator. They are then injected into the high-energy ring (HER) of the PEP-II storage rings in a clock-wise direction. Figure 1-1 shows the Linac and storage rings.

The electron beam in the Linac is then aimed at a tungsten target to produce positrons which are then accelerated to 3.1 GeV and injected into the low-energy ring (LER) in a counter-clockwise direction (Figure 1-2). The center-of-mass frame energy is tuned to make $\Upsilon(4S)$ particles which then decay to two B s. The unequal energies of the beams give a boost of $\beta\gamma$ =0.56 to the $\Upsilon(4S)$. This creates an average separation of the decay points of the two neutral B^0's of $\beta c\gamma\tau$ = 250 μm, which is crucial for studying CP violation modes.

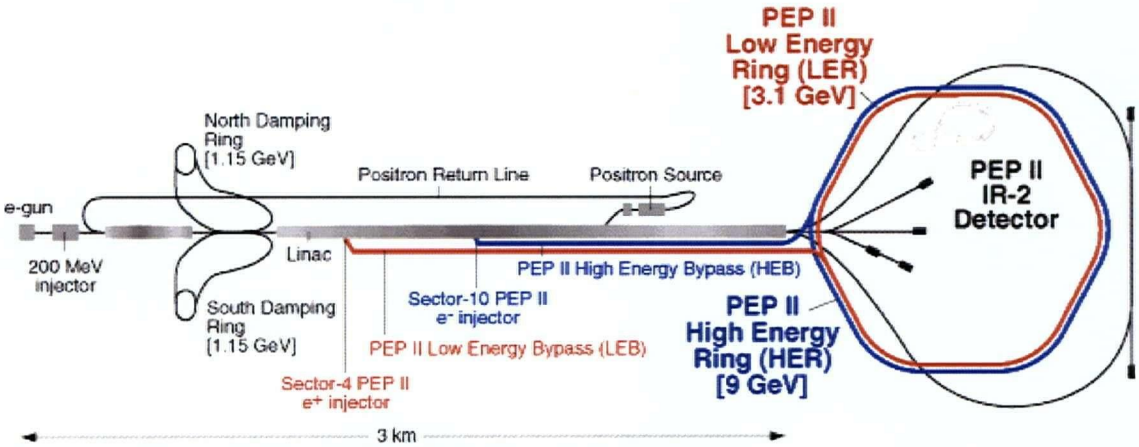

Figure 1-1: Schematic view of PEP-II LINAC and storage rings

1.3 The PEP-II Storage Rings

The PEP-II B factory is designed to operate at a center-of-mass energy (E_{CM}) of 10.58 GeV, corresponding to the $\Upsilon(4S)$ resonance. The designed luminosity for this energy is 3×10^{33} $cm^{-2}s^{-1}$.

The storage rings are stacked in a single 2200 meter PEP-II circular tunnel and cross at the interaction point (IP). The low-energy positron ring lies on top of the high-energy electron ring. The particles are kept in orbit by bending and focusing magnets and radio frequency acceleration. After the IP the beams are redirected back into their individual rings. The beams actually cross at two points: one is the IP and the other is a point 62 cm away from the IP. However, they do not collide at both points since they are being kept separated horizontally by a bending magnet at the second point. To ensure the maximum luminosity, the beams must be strongly focused near the IP. This is achieved by the placement of magnets very close to the IP in the interaction region (IR).

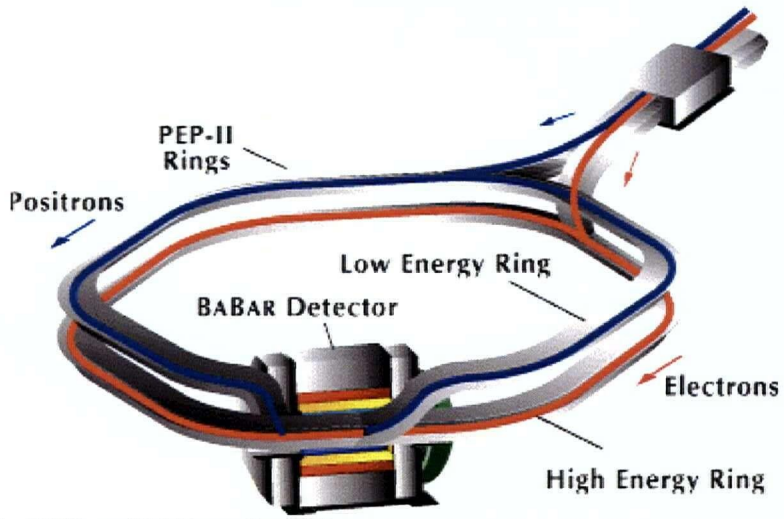

Figure 1-2: The HER and LER configuration.

1.4 The Interaction Region (IR)

Figure 1-3 shows the PEP-II IR in detail. As you can see, the positron beam enters from the lower right and exits through upper left. The electrons enter left and exit right. The positron and electron beams are called LEB (Low Energy Beam) and HEB (High Energy Beam) respectively. Note that the scale of this figure is exaggerated in the vertical direction. Also, the quadrupole magnet, Q1, and dipole magnet, B1, are inside the detector.

The first magnet encountered after the IP is B1, which horizontally separates the beams at the crossing point 62 cm downstream of the IP. The next two magnets, Q1 and Q2, are situated on the LER. Q1 focuses both the electrons and positrons in the horizontal direction.

The Q1 and B1 magnets are located inside a 1.5 T magnetic field and thus must be permanent or superconducting. On the other hand, space constraints do not permit B1 to be superconducting and therefore both Q1 and B1 are permanent, samarium cobalt magnets.

Q2 is situated three meters away from the IP and the HER to give horizontal focusing only to the LER.

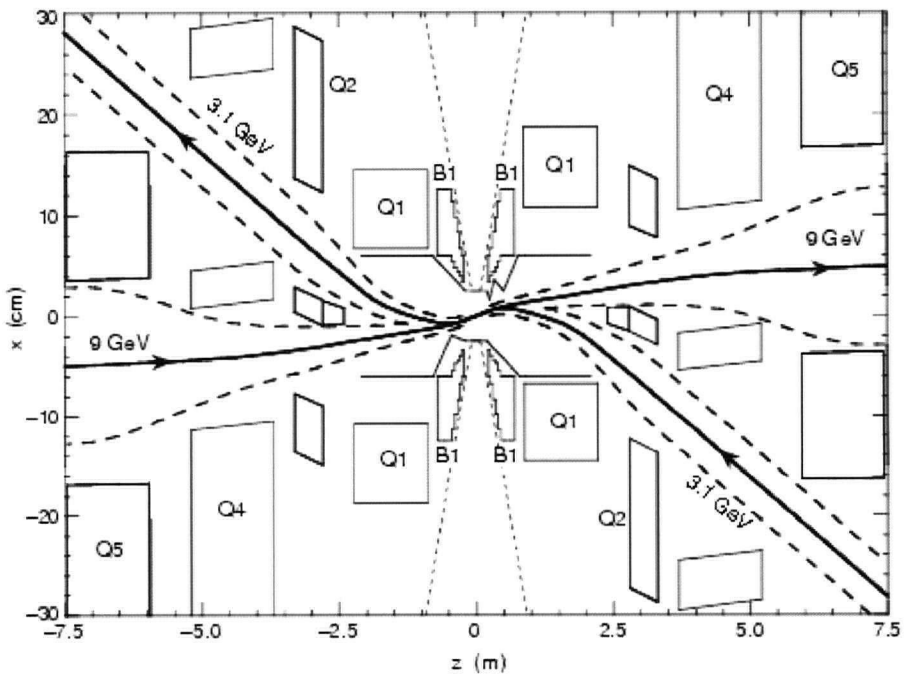

Figure 1-3: Plan view of the IR. The dashed lines represent the 300 mr detector acceptance cutoff.

There is also a support barrel inside the detector volume on which the Q1 and B1 magnets, the IP beam pipe, and the vertex detector are suspended (see Figure 1-4).

Bending the beams with the magnets near the IP is the source of a large amount of synchrotron radiation. To prevent this radiation from damaging the beam pipe, water-cooled beryllium masks are placed inside the B1 magnet.

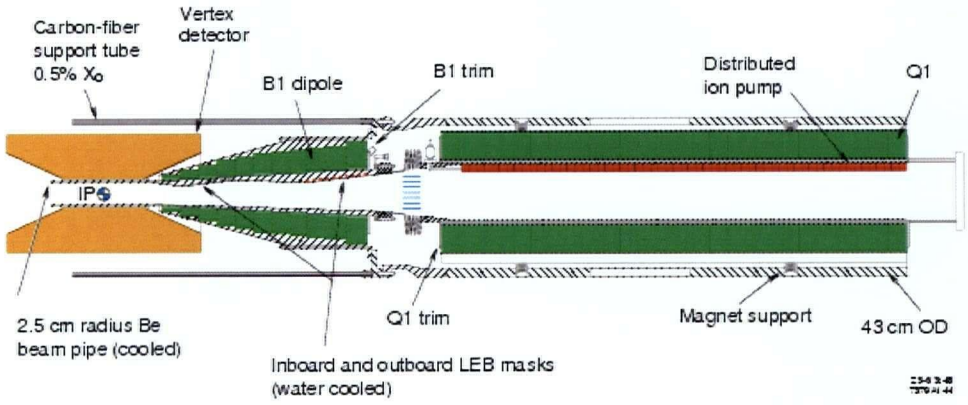

Figure 1-4: Support barrel for IR

1.5 The BaBar Detector

The BaBar detector has five major components that are concentric around the beam pipe and are offset in the forward direction. (Figure 1-5, Figure 1-6)

From inside to outside, the components are:

1. Silicon Vertex Tracker (SVT)
2. Drift Chamber (DCH)
3. Detector of Internally Reflected Cherenkov light (DIRC)
4. Cesium iodide Electromagnetic Calorimeter (EMC)
5. Instrumented Flux Return (IFR)

We will discuss these five components briefly in the following sections.

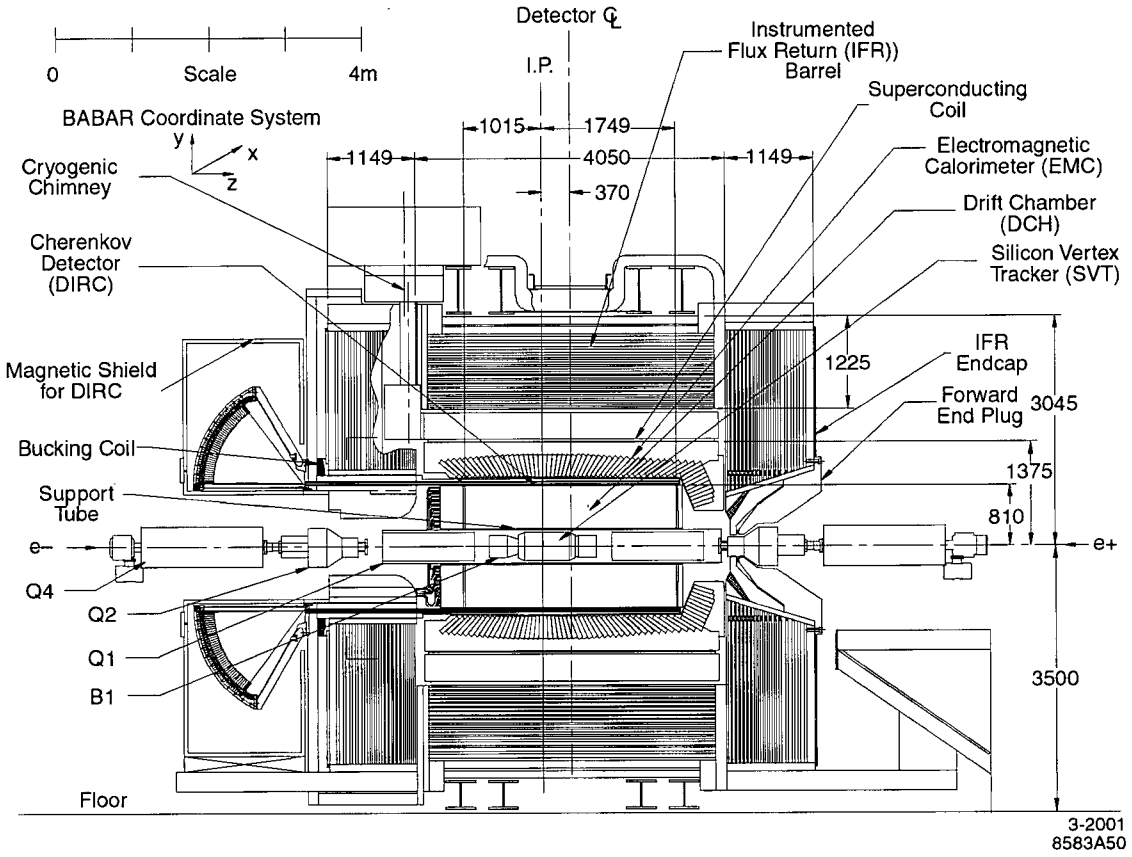

Figure 1-5: BaBar detector longitudinal section.

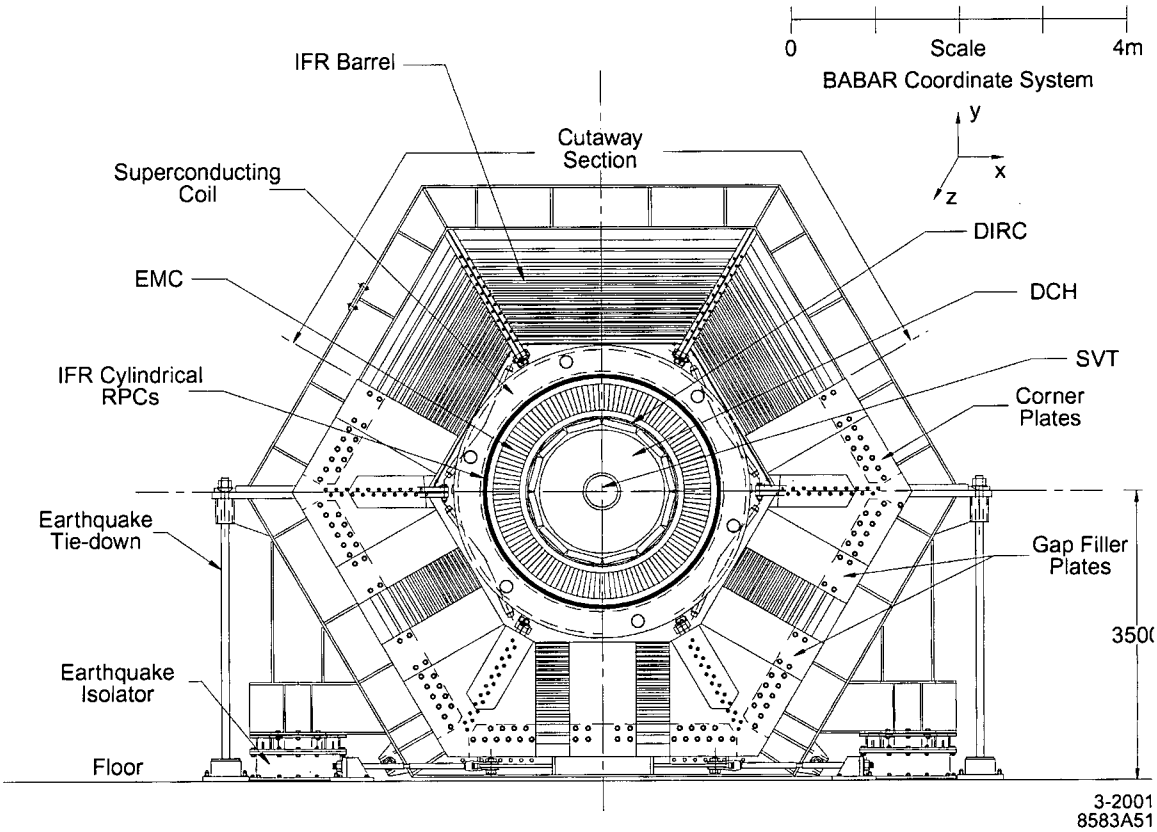

Figure 1-6: BaBar Detector end view.

1.5.1 The Silicon Vertex Tracker

The SVT is the innermost detector layer in BaBar. The main purpose of the SVT is to reconstruct the spatially separated decay vertices of the two primary B mesons in order to determine the time between the two decays. The basic physics requirement for the silicon tracker is the ability to measure the z separation between the two vertices to an accuracy of better than ~135 μm. This value corresponds to one half the mean separation between the two B mesons at PEP-II. This translates into a single vertex resolution of better than about 80 μm for the fully reconstructed B-decay. The silicon vertex tracker also measures the angle of charged particles. The SVT is composed of five double-sided layers of segmented silicon, running in strips in the z and φ directions[3]. Greater precision aids in pattern recognition, vertex reconstruction and background rejection, so the SVT is designed to have an intrinsic hit resolution of 10 – 15 μm for the inner layers and 30-40 μm for the outer ones. Figure 1-7 and Figure 1-8 demonstrate dimensional cutaway and cross section view of SVT.

[3] In BaBar the z-axis is defined to be in the direction of the beam. φ is defined to be the angle between the projection of the track on x-y plane and the x-axis ($0 \le \varphi \le 2\pi$). θ is defined to be the angle between the track and the z-axis ($0 \le \theta \le \pi$).

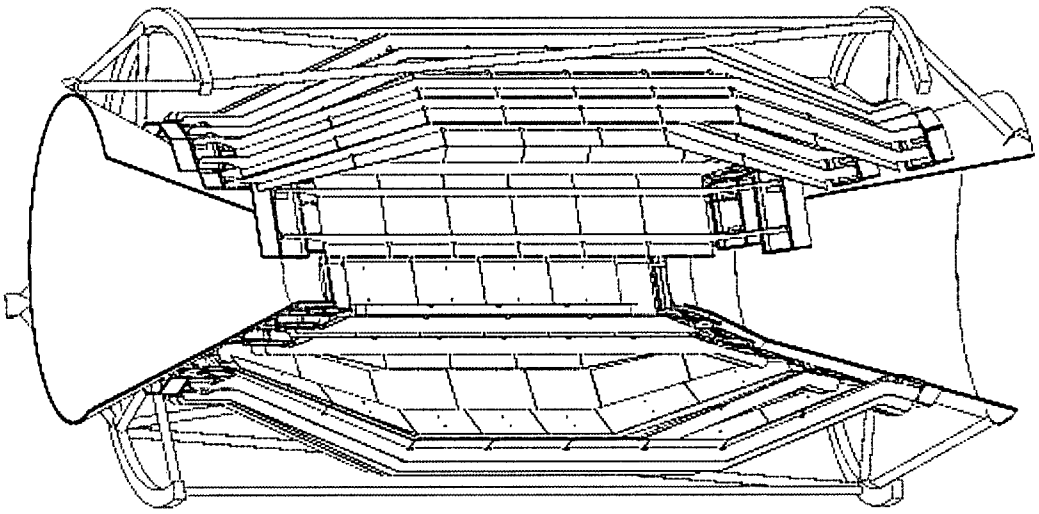

Figure 1-7: Three dimensional cutaway view of SVT.

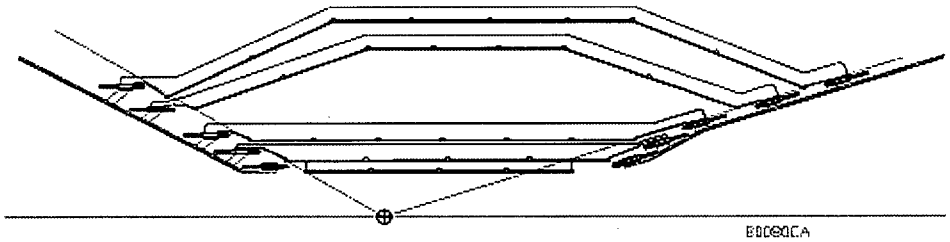

Figure 1-8: Cross section view of the SVT in a plane containing the beam axis.

1.5.2 The Drift Chamber

The drift chamber (DCH) surrounds the SVT and is the primary tracking device in the BaBar detector. The principal purpose of the drift chamber is the efficient detection of charged particles and the measurement of their momenta and angles with high precision. It also provides pattern recognition.

The reconstruction of exclusive CPV B decay modes[4] requires excellent momentum resolution. This is also required for the study of B meson mixing and many other analyses where the momenta are required when using leptons as tagging particles. The DCH provides tracking information for particles with momenta greater than 100 MeV/c. Since the average momentum of charged particles produced in B and D meson decays is less than 1 GeV/c, multiple scattering is a significant, if not the dominant limitation on the track parameter resolution. In order to reduce this

[4] Exclusive decays are the ones in which the final states are fully described.

contribution, material in front of and inside the chamber volume has to be minimized.

The DCH is also expected to provide particle identification using dE/dx measurements with a resolution of about 7% for low momentum tracks and a momentum resolution of $0.3 \times p_t$ for tracks with momentum greater than 1 GeV. The design also has to take into account high occupancies due to lost beam particles.

A schematic view of the BaBar drift chamber is shown in Figure 1-9. The DCH is a 280 cm long cylinder with inner radius of 23.6 cm and outer radius of 80.9 cm. Since the collision products are boosted forward in the lab frame, the detector is designed to reduce the amount of material in the forward direction.

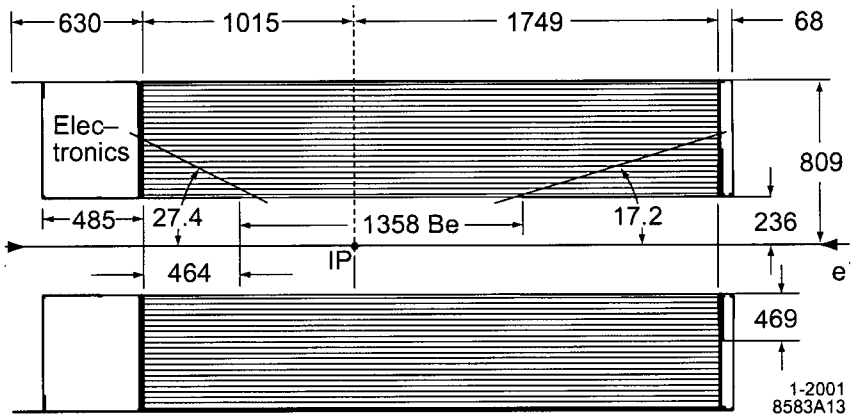

Figure 1-9: Longitudinal section of the DCH with principal dimensions; the chamber center is offset by 370mm from the interaction point (IP).

To accomplish this, the outer half of the forward endplate beyond the radius of 46.9 cm, is only 12 mm thick, compared to the inner half of the forward endcap and rear endplate, which have a thickness of 24 mm. All of the readout electronics and cooling apparatus are mounted on the rear plate. The inner cylinder is made of 1mm beryllium while the outer cylinder consists of 2 layers of carbon fiber on a Nomex core.

The chamber itself is made up of ten axial (A) and stereo (U,V) superlayers of 7104 hexagonal drift cells, each with a typical dimension of 1.2×1.8 cm^2. The superlayers in turn, are made up of four more layers and are set up in a AUVAUVAUVA configuration, Figure 1-10. For the stereo super layers, the stereo angle varies from 40 mr to 70 mr in the outermost layer.

The chamber is filled with a helium-isobutane (80%: 20%) gas mixture. The field wires are grounded and are 120 μm and 80 μm in diameter. The Guard wires are subject to a nominal voltage of 340 V from high voltage assemblies at the rear endplate. They both are made from gold-plated aluminum. The sense wires are 20 μm gold plated tungsten-rhenium and are subject to a potential of 1900-1960 V at the boundaries of the superlayers. High voltage assemblies mounted on the rear end plate also supply a nominal voltage of 900 V for the clearing wires, also made from gold plated aluminum, at the boundaries of the superlayers.

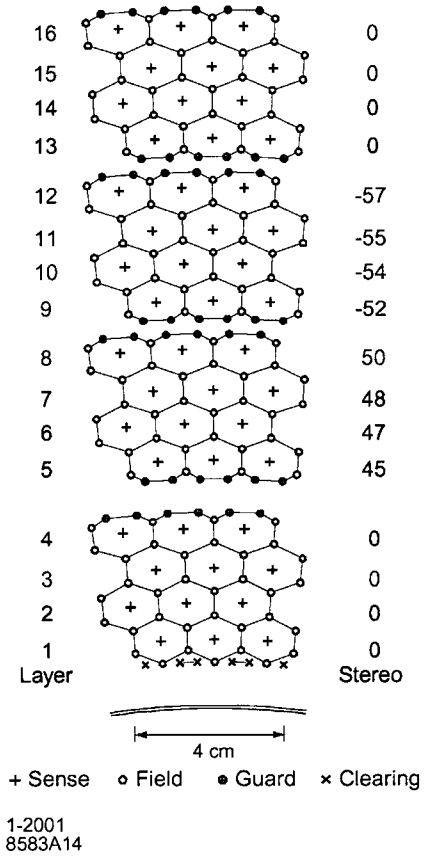

Figure 1-10: Schematic layout of drift cells for the four innermost superlayers. Lines have been added between field wires to aid in visualization of the cell boundaries. The numbers on the right side give the stereo angles (mrad) of sense wires in each layer. The 1 mm-thick beryllium inner wall is shown inside of the first layer.

1.5.3 Detector of Internally Reflected Cherenkov Radiation (DIRC)

The DIRC is a new type of Cherenkov-based detector, unique to BaBar, devoted to Particle Identification (PID). Specifically, it is designed to provide excellent kaon identification for both tagging purposes and at high momenta for rare B meson decays. It provides good π/K separation from 0.7 GeV to 4.2 GeV. To differentiate the two-body decay modes $B^0 \to \pi^+\pi^-$ and $B^0 \to K^+\pi^-$, the DIRC must be able to separate pions from kaons up to about 4 GeV at large dip angles in the laboratory frame.

Cherenkov photons are emitted by a particle traveling faster than the speed of light in the ambient medium. The Cherenkov angle (θ_C) is related to the index of refraction n and the speed of the charged particle $\beta = v/c$ through the relationship:

$$\cos(\theta_C) = \frac{1}{n\beta}$$

Equation 1-1

Therefore by measuring the angle of the Cherenkov light and the momentum of the track, the mass of the charged particle can be determined.

The high optical quality of the fused silica bar preserves the angle of the emitted Cherenkov light. An advantage of the DIRC for an asymmetric collider is that the high momentum tracks are boosted forward, which causes a much higher light yield than for particles at normal incidence. This is due to two effects: the longer path length in the fused silica and a larger fraction of the produced light being internally reflected in the bar.

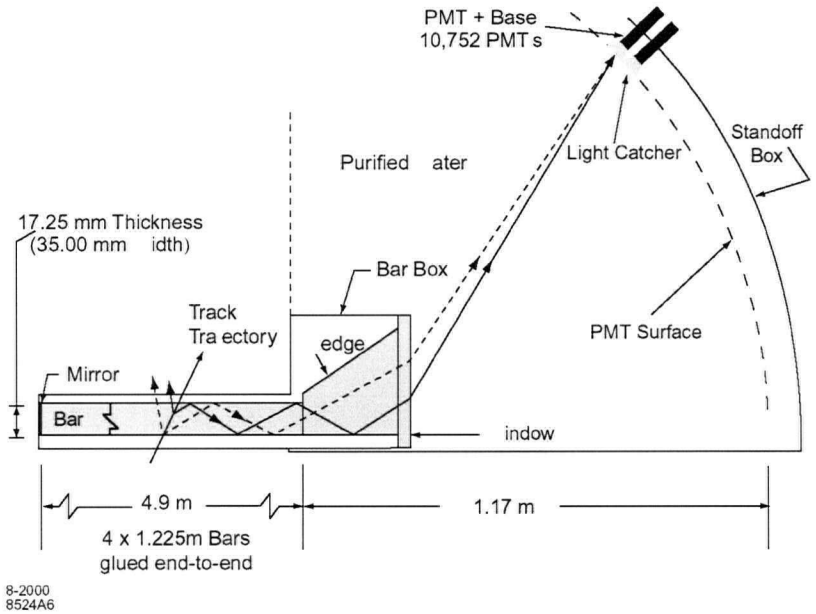

Figure 1-11: Schematics of the DIRC fused silica radiator bar and imaging region. Not shown is a 6mrad angle on the bottom surface of the wedge

The DIRC radiator is made up of 144 synthetic fused silica bars, which are 1.7cm thick, 3.5cm wide and 4.9m in length, Figure 1-12.

The fused silica bars have a mean refractive index of 1.474[5]. Cherenkov radiation produced in the fused silica bars are sent to the backward end of the detector by means of total internal reflections, Figure 1-13.

This ensures that the angle of production of the photon is preserved. Photons that are produced in the forward direction are reflected by mirrors and are sent to the back via the aforementioned method. At the backward end of the DIRC the image is expanded in 6 m^3 of purified water that has a refractive index comparable to that of fused silica in order to reduce the distortion due to refraction at the fused silica-water border. The

[5] The refractive index varies slightly with the wavelength. In general, the refractive index is some function of the frequency ν of the light, thus $n = n(\nu)$, or alternately, with respect to the wave's wavelength $n = n(\lambda)$.

Cherenkov light is detected by an array of 10752 Photomultiplier Tubes (PMTs) each with a diameter of 2.82 cm.

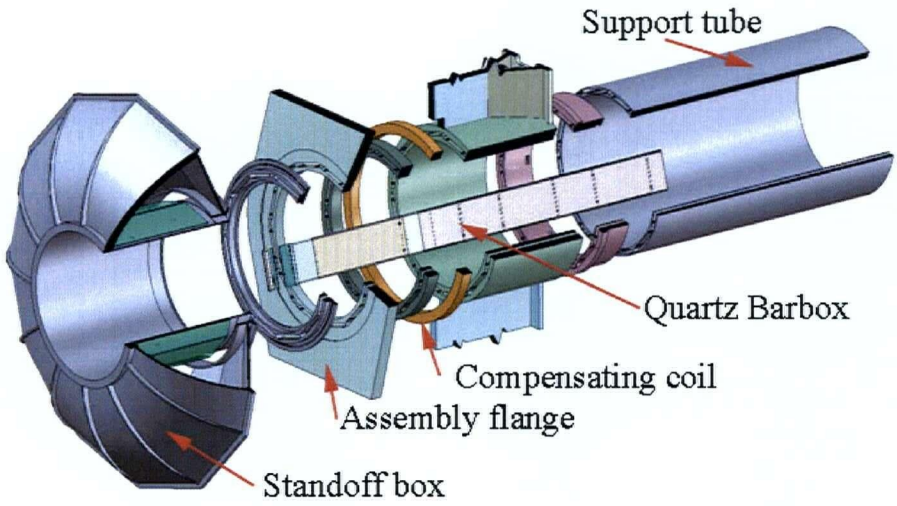

Figure 1-12: The DIRC

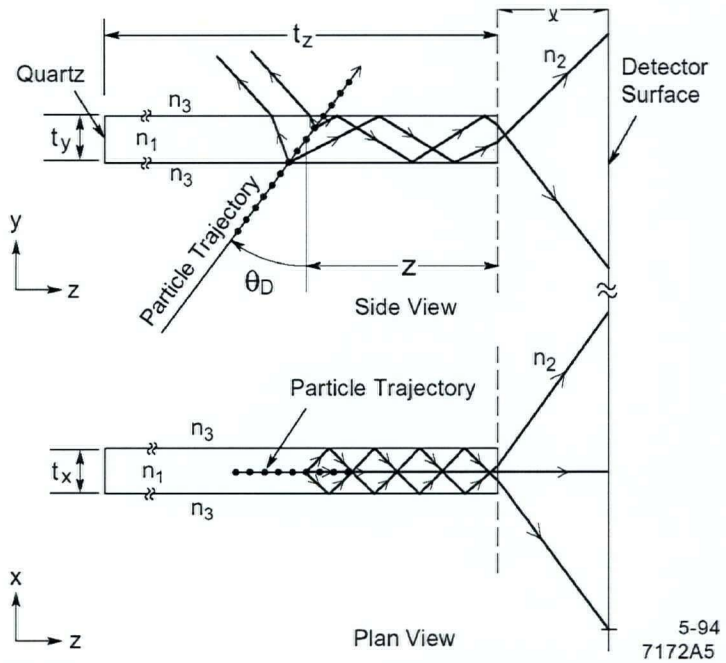

Figure 1-13: The Cherenkov radiation produced by charged particles traversing the fused silica is total internally reflected to the detector surface. In doing this, the production angle of the radiation is preserved.

The operating high voltage (HV) of the PMTs is ~1.14 kV, with a range between 0.9 and 1.3 kV. The PMTs operate directly in the water and are situated about 1.2 m away from the end of the fused silica bars. They detect Cherenkov radiation in the visible and near-UV spectrum. The PMTs and the water are contained in an enclosure known as the Stand Off Box (SOB), which is magnetically shielded to provide a suitable

environment for the PMTs. The SOB extends through the steel of the flux return in the backward end.

1.5.4 Electromagnetic Calorimeter (EMC)

Energy measurements are made by the EMC. It is designed to measure electromagnetic showers over the range of 20 MeV to 9 GeV with high efficiency, and excellent energy and angular resolution. This is necessary to allow the detection of photons from π^0 and η decays as well as from electromagnetic and radiative processes. The EMC is also important for particle identification of electrons and muons.

The BaBar EMC is made up of 6580 quasi-projective thallium-doped cesium iodide (CsI) crystals that cover a solid angle of $-0.78 \leq \cos(\theta) \leq 0.96$ in the laboratory frame. The front faces of crystals have a typical dimension of about 5cm.

The EMC consists of a cylindrical barrel section, offset in the forward direction, with a forward canonical endcap. The barrel has an inner radius of 90 cm and outer radius of 135.6 cm. The barrel is constructed of 250 μm thick carbon fiber composite compartments that house individual crystals. Each crystal is wrapped with a diffuse reflecting material on its sides and a reflector on its front face.

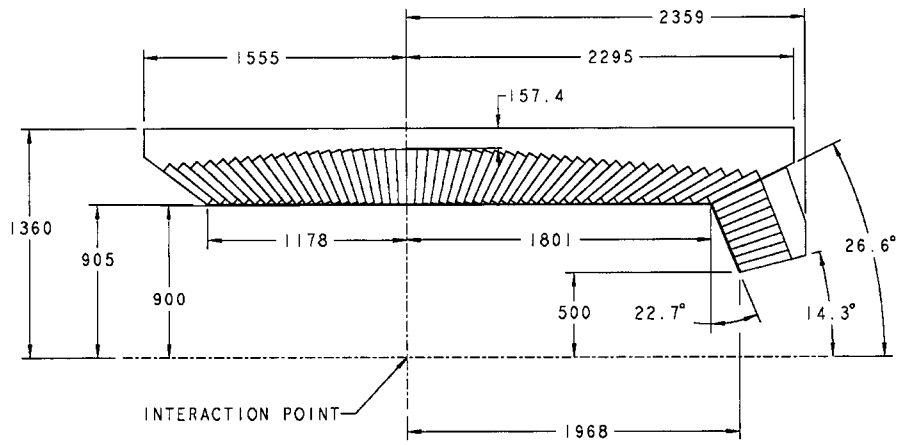

Figure 1-14: A longitudinal cross section of the EMC (only the top half is shown) indicating the arrangement of the 56 crystal rings. The detector is axially symmetric around the z-axis. All dimensions are given in mm.

The crystals are tapered along their length with trapezoidal cross sections. The average area of the front faces of the crystals is 4.8×4.7 cm, while the back face area is 6.1×6.0 cm. They vary in length in $0.5X_0$ steps from $17.5X_0$ in the forward part of the barrel to $16.0X_0$ in the backward part, with $17.5X_0$ in the forward endcap, where X_0 is a radiation length of the cesium iodide. The barrel and endcap have total crystal volumes of 5.2 m^3 and 0.7 m^3, respectively.

Photodiodes and preamplifier packages provide readout and are located at the outer end of each compartment. The choice of photodiodes as readout devices was made

since readout must be accomplished in the 1.5 T magnetic field. The forward endcap is segmented vertically into two pieces, each of which can be removed separately allowing relatively easy access to the barrel end region. Figure 1-14 shows a cross-sectional view of the EMC.

The CsI crystals chosen for the EMC ensure the necessary high resolution and efficiency. The primary concerns are minimizing the cost and the amount of material in front of the calorimeter. In the interest of cost, the size of all internal components has been restricted and there is no backward endcap. Material between the IP and the calorimeter has been minimized in all internal components. The largest contributor to radiation lengths in front of the calorimeter is the DIRC. The total amount of material including the beampipe, vertex tracker, drift chamber and DIRC represents ~0.23 X_0 at normal incidence. The major source of material in front of the endcap is the DCH endcap.

1.5.5 Instrumented Flux Return (IFR)

The IFR provides muon identification and neutral hadron detection. It makes use of the large iron structure needed as the magnet return yoke. The iron is segmented and instrumented with Resistive Plate Counters (RPCs) and consists of a central part (Barrel) and two plugs (End Caps), Figure 1-15.

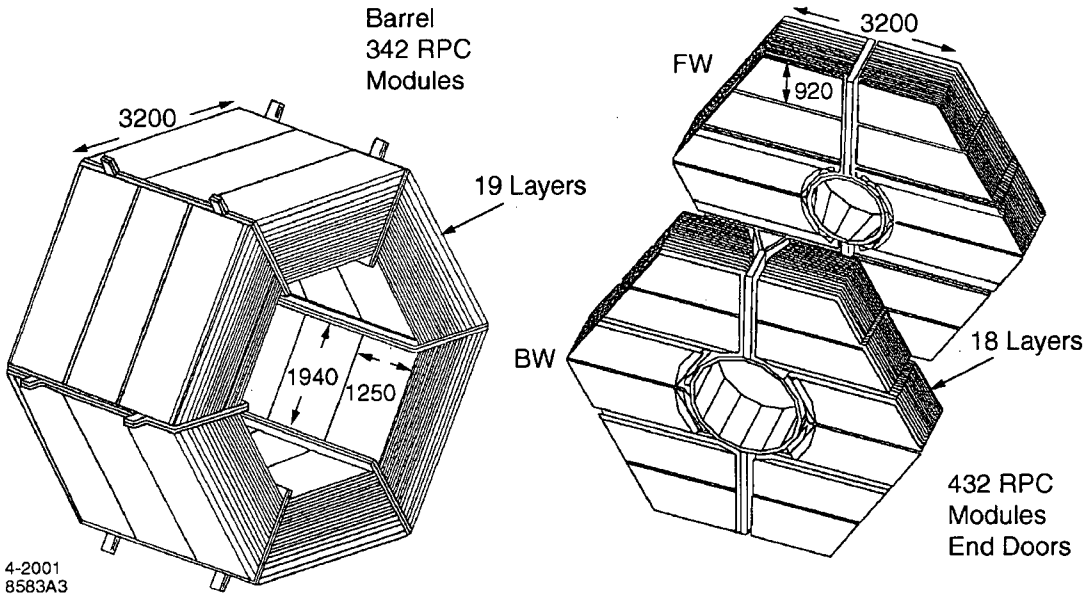

Figure 1-15: Overview of the IFR: Barrel sectors and forward (FW) and backward (BW) end doors; the shape of the RPC modules and their dimensions are indicated.

One of the interesting features of the BaBar experiment is the graded segmentation of the iron, which varies from 2 – 10 cm, increasing with the radial distance from the interaction region. The gap between plates is 3.5 cm in the inner layers of the barrel and 3.2 cm elsewhere. This segmentation is due to detailed Monte Carlo studies which have

shown that low-momentum muon identification and K_L detection improves for a given amount of absorber as the thickness of iron plates decreases. This effect is most important in the first absorption length, so that grading the segmentation improves the performance without increasing the number of layers significantly. Planar RPCs are inserted before and after each plate making a total of 19 RPC layers in the barrel with a total thickness of 65 cm. The endcaps have 18 RPC layers of a total thickness of 60 cm. Each endcap door has a hexagonal shape and is divided vertically into 2 halves to allow access to the inner detectors. Additionally, two layers of cylindrical RPCs are installed between the EMC and the magnet cryostat to detect particles exiting the EMC. There are a total of 806 RPC modules, 57 in each of the six barrel sectors, 108 in each of the four half-end doors, and 32 in the two cylindrical layers.

The planar RPCs are made of two Bakelite sheets, 2 mm thick. These are separated by a gap of 2 mm which is filled with a gas mixture of 56.7% argon, 38.8% freon, and 4.5% isobutene. The inner surfaces are coated with graphite and connected to high voltage (~8 kV) and ground, and protected by an insulating mylar film. The RPCs detect streamers from ionizing particles via capacitive readout strips. The width of the strips varies between 16.0 and 38.5 mm depending on the location of the RPC.

The main goal of the IFR detector is to achieve the highest practical muon tagging efficiency. About 18% of all B decays contain at least one muon in the region covered by the BaBar detector. Charged tracks found in the DCH will be matched with segments in the IFR and be identified using a detailed analysis of the hit patterns in the active detectors. Neutral hadron (K_L and neutrons) identification will result from a detailed analysis of the hit patterns in the active detectors with no track matches from the tracking chambers.

Due to unforeseen problems with the RPCs, they have been gradually failing and suffering from performance degradation during the first 6 years of operation. In the summer of 2004 and 2005 the RPCs will be replaced with Limited Streamer Tubes (LSTs) in order to restore the efficiency for muon identification to the highest possible level.

2

The Data Acquisition and BaBar Software

The previous chapter described the PEP-II collider used to produce $B\overline{B}$ events, and the BaBar detector used to record them. The raw data recorded by the detector goes through many stages of processing before it shows up as histograms in a published report or thesis. The initial filtering and reconstruction of the data are done centrally on BaBar computers at SLAC.

The reconstructed data is stored in BaBar's central database, the Event Store, where users can access and use it to produce their own ntuples for offline analysis. This chapter describes in more detail how data is processed at BaBar. For more information about this subject you can refer to [6].

2.1 Data acquisition (DAQ)

The data acquisition (DAQ)[8]system's job is to get data from the detector to the online prompt reconstruction (OPR) system. There are two main stages of processing that the data must undergo before it can be passed to OPR. First, it must be digitized or translated from raw detector signals to "computer language". Second, it must be filtered by the trigger system to get rid of as much background as possible so that OPR can cope with the rate of incoming data. The main steps in DAQ are the following:

- **Front-end electronics (FEEs):** For initial processing each subdetector has its own front-end electronics (FEEs), which are located right on the subdetector and receive the raw detector signals. The FEEs amplify and digitize the signals, and perform some other subdetector-specific processing. They then send the data to the subdetector's readout modules.

• **Readout modules (ROMs):** For primitive construction each subdetector also has its own readout modules (ROMs), located on the detector. The ROMs use the FEE data from the EMC and DCH to construct basic data objects called primitives and send them to the Level-1 trigger.

• **Level-1 trigger**: The Level-1 (L1) trigger is a quick hardware trigger that filters out the most "obvious" background. It uses the primitives from the DCH and EMC ROMs and other simple selection criteria to make a quick decision whether to accept or reject events.

• **Feature extraction**: If an event passes the L1 trigger, a signal is sent to the ROMs telling them to collect the rest of the FEE data. The readout modules then perform feature extraction, transforming the raw data into useful information such as particle hit time and energy. The output of feature extraction is called an event fragment.

• **Event assembly**: The DAQ system collects the event fragments from the ROMs of all the subdetectors, and assembles them into an event for the Level-3 trigger.

• **Level-3 trigger:** The Level-3 trigger performs the final event selection. It runs in software, and uses more complex algorithms than the Level-1 trigger to decide whether to accept or reject events. Events passing the L3 trigger are sent to the online prompt reconstruction (OPR)[9] system, which performs the main event analysis.

2.2 The Trigger

A trigger is a very important part of high-energy particle physics experiments, particularly those with very high backgrounds. The trigger's job is to act as a filter, quickly selecting interesting physics events for further analysis while rejecting background. Since it must keep up with the event rate, the trigger's selection criteria are typically very simple, based on things like event topology or track multiplicity. Many experiments are limited by deadtime - the time when data is not being recorded because the data-taking electronics can't keep up with the event rate. A good trigger should minimize or eliminate deadtime. For experiments with particularly high rates, this usually requires a multilevel trigger, with each level receiving data at a lower rate and using more complex selection criteria than the previous level. The complete reconstruction of an event by the online prompt reconstruction (OPR) system is a complicated process, far more complex than the loose pseudo-reconstruction performed by the trigger system. The BaBar OPR system can accept events at a maximum rate of about 120 Hz. This is fast enough to allow for the complete reconstruction of all interesting events at BaBar - $B\overline{B}$ physics, charm physics, τ physics, and two-photon physics events - occur at rates of several Hz each, for an overall rate of less than 20 Hz.

The detector records not only the events of interest, however, but also background events - bhabha scattering, beam-induced backgrounds from the interactions of "lost" beam particles, and cosmic ray muon events. These events are not "interesting" in the sense that they are already well understood, and there are no BaBar projects that study them (although they may be used for many calibration measurements). Due to its high luminosity, PEP-II has very high background rates for an e^+e^- experiment - at the design luminosity of $3\times10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ (which BaBar has exceeded), the background rate is about 1200 Hz. Even if BaBar wanted to log all these background events (which it doesn't), it is beyond OPR's capacity. BaBar's trigger system must filter out most of this background in order to bring the incoming data rate down to a level that OPR can cope with.

BaBar's Trigger has two levels, the Level 1 (L1) hardware trigger and the Level 3 (L3) software trigger. The purpose of the L1 trigger is solely to reduce backgrounds while remaining as "open" as possible to events of interest. The L3 trigger then uses the DCH and EMC information to select the physics events of interest to be sent for reconstruction.

The maximum permissible rates for the L1 and L3 triggers are set by the maximum rate of the next system downstream. For the L1 trigger, this is the data acquisition (DAQ) system, which has a maximum rate of 2 kHz. For the L3 trigger, it is the online prompt reconstruction (OPR) system, which can reconstruct events at a maximum rate of 120 Hz.

2.2.1 The Level 1 (L1) trigger

The Level-1 trigger receives the raw data straight from the detector, before any processing. It consists of a drift chamber trigger (DCT), an electromagnetic calorimeter trigger (EMT), and a global trigger (GLT)[6]. The EMT and DCT construct basic data objects called primitives from the raw hits in the subdetectors. The idea is that a primitive corresponds to a (possible) particle. The DCT primitives are long and short tracks in the DCH. The EMT primitives are clusters of crystals with energy above a certain threshold. The results from both the EMT and DCT are sent to the GLT, which decides - mostly on the basis of track and cluster multiplicities and topologies- whether to reject the event or allow it to be sent to Level 3.

2.2.2 The Level 3 (L3) trigger

Once the events have been assembled, there is plenty of information that can be used to filter out background events. The Level 3 trigger runs in software, and uses more complex algorithms to analyze the event data, combining DCT tracks and EMT clusters

[6] There is also an IFR trigger (IFT), but it is used only for diagnostic purposes.

from L1 with the full DCH and EMC information. Like L1, the L3 trigger has separate and orthogonal DCH and EMC triggers[7]. The L3 trigger passes various sets of *trigger lines*[10]. The physics events of interest to most users are in the physics line. However, L3 also passes diagnostic and calibration lines for diagnostic and calibration studies.

Events passing the L3 trigger are written to temporary files, called XTC (eXtended Tagged Container) files, which serve as the input to OPR. Each XTC file contains all the events taken for a single run[8] of the collider, typically about 300,000 events.

2.3 BaBar software

BaBar software is written in a computer language called C++, which is designed to support object-oriented programming[11]. Object-oriented programming makes it easy to model real-life objects and concepts by designing classes to represent different types of objects. This is useful for BaBar because it allows for the association of things like particles and detectors with software objects. For example, the class *ChargedTracks* was created to represent charged tracks in the DCH and SVT. Furthermore, C++ allows for a hierarchy of classes, via the mechanism of inheritance - for instance, *ChargedTracks* inherits from the generic particle candidate class *BtaMicroCandidate*.

BaBar software for reconstruction, simulation, and ntuple production is organized in terms of packages, self-contained sets of code intended to perform a specific task (for example, to find calorimeter clusters). BaBar is constantly improving and updating its software, with regular releases of the most recent stable version of each package, as well as the libraries and binaries needed for particular machine architectures. Most researchers use the release designated as the current release; this is the most recent release to have reached an acceptable level of quality.

[7] This means that the triggers are independent; for example, if the EMC has problems, the efficiency of the DCH triggers does not change.

[8] A run is the time between two beam injections at PEP-II. A single run is about 1 hour.

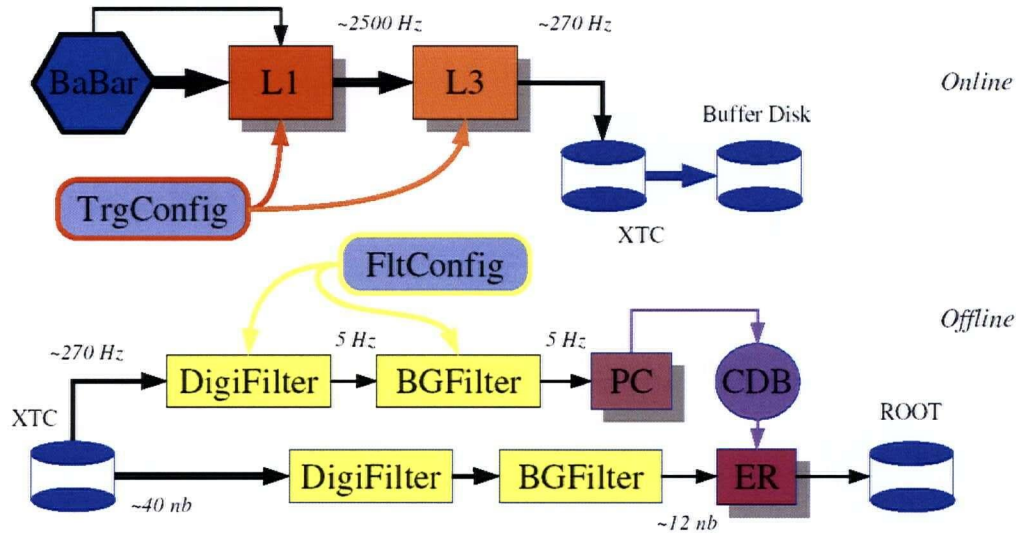

Figure 2-1: This figure demonstrates the levels of filtering an event passes before it is stored for an offline study.

Releases are numbered chronologically, so that the highest numbers represent the most recent releases. Releases with the same first number X are very similar, and are often referred to as "the release X series," or just "release X". For example, the release 10 series includes releases 10.2.3b, 10.3.0f, 10.4.1a, and far too many more to list here. The differences between the releases in a series are just minor bug fixes and the like. On the other hand, releases with different first numbers are very different - data processed using a release in the release 8 series has a different format than data processed using release 10. Since the most recent releases represent the latest and greatest software, BaBar implements a system of reprocessing the data - completely redoing the reconstruction for all old, already-reconstructed data, but this time using the new release. Thus, you have the option of analyzing data initially processed using release 8, with the new and improved release 10 software; but not vice versa. A given analysis should use compatible releases of the different types of software. In practice what this means is that the releases used for reconstruction, simulation, and ntuple production should have the same first number. For example, the data set from Runs 1, 2 and 3 were reconstructed using release 12 reconstruction software, so for these Runs I used release 12 ntuple production software to make my real data ntuples, and I used SP5 ("Simulation Production 5") software, which is also release 12, to make my simulated ("Monte Carlo") ntuples. Using compatible releases ensures that all the code works together properly and also that the background conditions are the same in the real data and Monte Carlo simulated sets.

2.4 Online Prompt Reconstruction (OPR)

Reconstruction is the last stage of data processing to be performed centrally at BaBar. The aim is to reconstruct particle candidates from the raw hits in the detector.

This thesis has already described the first few steps of this process, with the detector hits being used to form primitives for the trigger, and assembled into events for the L3 trigger. But it is only after an event is selected by the L3 trigger and passed to the online prompt reconstruction (OPR) system [9] that the true reconstruction begins. The OPR system aims to completely reconstruct all physics events passed by the Level-3 trigger within several hours of acquisition (hence "online" and "prompt"). The need to process events quickly comes not only from the desire to obtain the latest physics results as quickly as possible, but also from the need to provide feedback for the detector operations staff so that they can fix problems as they arise. This was particularly important in the first stages of the experiment. The rate of physics passed by L3 trigger is about 100Hz most of which is background[9]. The OPR system was designed to be able to accept data at rates of up to 120Hz (with zero deadtime), and to finish processing it within several hours (minimizing latency).

Calibrating the detector is a continuous and important part of detector operations. BaBar uses a system of rolling calibrations, in which calibration information in the Conditions Database is continually updated and used in reconstructing events from the same time period. These conditions are also used in the production and reconstruction of simulated Monte Carlo data.

2.5 Reconstruction algorithms

To translate the raw detector hits into a description of particles and their decays, the OPR system uses reconstruction algorithms implemented in software. Reconstruction takes place in three steps. First, the hits are reconstructed into the basic objects corresponding to individual particle candidates: tracks in the tracking devices, and clusters in the calorimeters. Second, particle identification (PID) algorithms are used to assign an identity hypothesis to each particle candidate. Finally, tagging creates a database of tag bits, simple Boolean or Boolean-like expressions for quick data skims. This section describes these steps in more detail.

2.5.1 OPR filters

There are two stages of filtering at the OPR level [10]. The *DigiFilter* is based only on the L1 and L3 trigger output, and is run before the full reconstruction. The main purpose is to select the physics line of physics events, and reject the diagnostic and calibration lines.

After some initial reconstruction another filter, *BGFilter* (background filter)[10], is applied. *BGFilter* consists of selectors that tag physics events as multihadron events, τ events, two-photon events, and so on. An event must have at least one of these tags set

[9] The rate of interesting physics is about 15Hz but L3 trigger can't do any better at getting rid of the remaining 85Hz of junk.

to true in order to be written to the Event Store database; otherwise it will be available only in the XTC file. Events passing *BGFilter* are collectively called *isPhysicsEvents* events.

In this analysis, I have generated my ntuple using only events that pass the *BGFPhiGamma* filter. The physics conditions for an event passing this flag are:

- The number of charged tracks should be less than or equal to 5 and greater than or equal to 2.
- One Isolated Photon with energy less than or equal to 6 GeV and greater than or equal to 4 GeV.
- All tracks should satisfy the relation $\cos(\varphi_{Trk} - \varphi_{Photon}) \leq 0.8$.
- Existence of at least 2 tracks in opposite hemispheres with an invariant mass consistent with $K_S \rightarrow \pi^+\pi^-$ or $\varphi \rightarrow K^+K^-$ (The latter is what I'm looking for as my signal event).

2.5.2 Track and cluster finding

The aim of the BaBar detector is to detect and identify particles; but the raw data obtained by the detector is in the form of hits in the various subdetectors, not particles or particle candidates. The first step in reconstruction is therefore to run algorithms which use the hits to form *basic objects* corresponding to particle candidates. In the tracking devices, the basic object is a charged track, formed by "connecting the dots" between the hits in the many layers of the DCH and SVT. In the calorimeters, where particles are identified based on their showers, the basic object is a *cluster*, a bunch of hits in the same general region of the calorimeter. Because of the magnetic field, the charged tracks in the SVT and DCH are helical. They would be exact helices if not for multiple scattering, energy loss in the material, and inhomogeneities in the magnetic field. The tracking algorithms are designed to look for ways to join the detector hits into near-helical tracks, taking these three effects into account. Track reconstruction begins in the DCH with the tracks and the event time (T0) estimate from the L3 trigger. The algorithm looks for track segments in the individual DCH superlayers, which consist of 4 layers each to allow for 3-out-4 majority logic in segment finding and also in triggering decisions. The algorithm then tries to piece the segments together to form tracks. The tracks are fit to the expected near-helix, and then the tracks from the DCH are extrapolated to the SVT. Finally, additional algorithms are run on the SVT hits to check for tracks that don't extend through to the DCH (corresponding to typically low momentum and low curve charged particles that decayed before they reached the DCH).

A particle showering in the EMC will typically deposit energy in several crystals. Groups of crystals containing the energy deposit from a single particle are called clusters or bumps. More specifically, a cluster is a set of adjacent crystals with the sum of their energies above a certain minimum energy. A cluster represents a single particle candidate, except in cases where two or more particles deposit energy in the same region. In these cases the clusters typically contain bumps - local energy maxima in the

clusters - and it is the bumps that represent single particle candidates. So the EMC reconstruction algorithms look for clusters, and then check the clusters for bumps. The particle candidates in the IFR are also called clusters, but as the IFR does not provide energy measurements. An IFR cluster is just a group of adjacent hits.

The last step is track-cluster matching. This is a search for tracks and clusters that correspond to the same particle candidate. The search is performed by extrapolating the charged tracks through the EMC and DCH and checking for any clusters along these extrapolated trajectories. Clusters that can be associated with charged tracks correspond to charged particles, and are often called charged clusters. Clusters that have no associated track represent neutral particles and are called neutral clusters. Thus, track and cluster finding outputs lists of charged and neutral particle candidates, where charged candidates are track-cluster associations or just tracks, and neutral candidates are clusters that don't have associated tracks.

2.5.3 Particle identification (PID)

Once the track and cluster particle candidates have been found, the next step is to determine which type of particles they represent. The only particles that are observed directly in the detector - that is, that form tracks and/or clusters - are electrons, muons, pions, kaons (K^+, K^-, K_L), photons, and protons. Other particles decay too quickly to be observed, so their existence and properties must be inferred from their track or cluster decay products. So the aim of particle identification is to identify each charged track and cluster as one of these six species of observable particle.

A particle is identified based on its mass and its quantum numbers; or equivalently, its mass and how it interacts with the material of the detector. Charged particles are much easier to identify than neutral particles, since both their charges and momenta can be determined from the curvature of their tracks, and their masses can be calculated using the momentum and either a velocity or energy measurement. At BaBar, charged particle energy (specifically, the ionization energy loss dE/dx) is measured in the SVT and DCH, and charged particle velocity can be obtained from a Cherenkov angle (θ_c) measurement in the DIRC. Neutral particles, on the other hand, must be identified based solely on their showers in the calorimeters.

At BaBar, particle identification is implemented via particle identification selectors, sets of cuts developed and maintained by the BaBar Particle Identification Group. There is a selector for each type of observed particle, and each has different levels of selection.

These are usually called "very loose", "loose", "tight" and "very tight" with looser cuts having higher efficiency but also higher misidentification rates. The selector algorithms run over the tracks and clusters, and assign tags to each one that indicates which selectors the candidate passed. For example, a candidate closely resembling a muon would pass the "tight" or "very tight" muon selector and would get a *muMicroTight* or *muMicroVTight* tag.

Here is a brief overview of how particles are identified at BaBar:

• *Electrons* are charged clusters that shower in the EMC with a characteristic electromagnetic shower shape, as parameterized by shower-shape variables like LAT (lateral moment of energy deposit).

• *Muons* are charged tracks that do not shower, and penetrate further in the IFR than any other particle.

• *Photons* are neutral clusters that shower in the EMC with a characteristic electromagnetic shower shape, as parameterized by shower-shape variables like LAT.

• *Charged pions, kaons, and protons* are identified and distinguished from each other (and from electrons and muons) using likelihood selectors. The discriminating variables are the ionization energy loss dE/dx in the SVT and DCH; and at high enough momentum, the Cherenkov angle θ_c and the number of photons N from the DIRC.

• *Neutral pions* (π^0) are neutral clusters that decay to two photons, $\pi^0 \rightarrow \gamma\gamma$.

• *Neutral long-lived kaons* K_L^0 are neutral clusters that are not neutral pions, that reach the IFR but don't penetrate very far, and whose EMC and IFR showers are characteristic of a hadron. (Neutral short-lived kaons K_S^0 are too short-lived to be directly observed and are reconstructed by their decay to $\pi^+\pi^-$ or $\pi^0\pi^0$)

The details for each selector can be obtained from the Particle Identification Group[13]. The output from reconstruction goes to BaBar's central database, the Event Store. The output of track and cluster finding and PID is a list of particle candidates; class *BtaMicroCandidate* in BaBar's C++ notation. The *BtaMicroCandidates* are sorted into more specific lists such as *ChargedTracks*, *CalorNeutral* and *GoodTracksLoose*. These fundamental lists are used by all BaBar collaborators for their analyses, and are stored at the Micro level of the Event Store.

2.5.4 Event tagging

The last step in reconstruction is called tagging. Event tags allow events of interest to be selected very quickly. Events are tagged with *tag bits*, which contain global information about events such as event parameters, which triggers and filters the event passed and interesting physics quantities. The tag bits are stored at the Nano level of the Event Store. Tag bits are used to create subsets of data corresponding to specific physics processes. These subsets are called *skims,* or *streams*.

Thus, OPR fills BaBar's Event Store database with particle candidates. The lists of *BtaMicroCandidates* are stored at the Micro level, and the tag bits are stored at the Nano level. These two databases contain all of the information needed for a typical

physics analysis.

2.6 Skims, streams, and collections

The Micro database contains many different types of interesting physics events. However, the typical researcher is interested in only one or a few types of event. For example, a researcher in BaBar's τ physics group will generally not need to study charmonium events. For this reason, each Analysis Working Group (AWG) defines its own selection criteria to produce its own *skims* or *streams* - subsets of the data, which contain the events of interest to that AWG. The selection criteria typically involve one or several tag bits in the Nano database. For example, the Charmonium AWG defines a skim called *Jpsitoll*, which picks out events with the *JpsiELoose* or *PsiELoose* or *JpsiMuLoose* or *Psi2SMuLoose* tag bits. (In case you can't tell from the names, the purpose of this skim is to select $B \to J/\psi\, X, J/\psi \to l^+l^-$ and $B \to \psi(2S), \psi(2S) \to l^+l^-$ events). For releases 10 and 12 the skim selections were run as part of the reconstruction. In release 10 there are 21 streams; for release 12 there are just 4 big streams.

The data subsets produced in skims and streams are called *collections*. Each collection is assigned a name that specifies all the information about the collection: the stream or skim name, the software release version used to produce it, its run number, and which version it is (there is more than one version if the collection has been reprocessed).

2.7 Ntuple production

Ntuple production involves running BaBar code to access collections in BaBar's Event Store, and storing some of the event information in a very convenient and portable data format called an *ntuple*.

The ntuples serve as a user's private copy of the data, so there is no longer any need to interact with BaBar Central Computing. The user can continue to work online with BaBar software if she desires; or she can copy the ntuples to her own computer and analyzes them offline. Either way, once the ntuples are produced the user can begin independent analysis.

To produce ntuples, the user needs two things: data from the Event Store and an executable to fill ntuples. To create the executable, the user needs BaBar ntuple production software. First the user checks out a release of ntuple production code and adds the appropriate packages. Once she has a package-filled release, the user sets the parameters for the preselection, the loose selection applied during the production of the ntuples. This involves making minor modifications to the code, such as the selection of particular tag bits, and the definition of the reconstruction criteria of the decay of

interest to the user. The preselection is distinct from the final selection, which the user imposes at the ntuple level during her independent, offline analysis.

Once the code is ready, the user links and compiles it to make the executable[10]. The executable will make ntuples, but to do this it needs data, so the user must communicate with the Event Store using a computer language called tcl ("tool command language"). This step is made somewhat easier by a BaBar tool called *skimData*, which can automatically generate tcl files to access particular collections. With the executable and tcl files listing the data, the user is ready to run the job. The executable reads in the data and writes out ntuples.

The ntuple production code does two main things: first, it performs additional reconstruction; and second, it stores the results in the ntuple format. As we've seen, Micro data consists of a list of *BtaMicroCandidates*, the C++ version of tracks and clusters. The ntuple production takes these lists and determines the most probable identities and decay trees of the particle candidates. Some of the most important reconstruction tasks in a typical ntuple production job are the following:

- **Nano tag filters -** Ntuple production usually begins with a pass through the Nano database of tag bits. As explained in Section 2.5.4, tag bits are used for quick filtering to find the events of most interest to the user.

- **PID decisions -** In Micro data, particle candidates are stored in lists of *BtaMicroCandidates*, and for each *BtaMicroCandidate* there are PID tags indicating which PID selectors the candidate passed. The ntuple production code uses this information to determine the most probable identity for each particle. It takes the lists of *BtaMicroCandidates* and separates them into lists of electron candidates, muon candidates, kaon candidates, and so on.

- **Composition -** Composition involves recreating the "decay trees" in the event and determining the mother and daughters of each particle.

- **Vertexing -** Vertexing takes decay trees and determines the most likely position of the vertex, and the most likely momenta of the particles at the vertex.

- **Kinematic fitting -** Kinematic fitting uses kinematic constraints to determine the best values for the momenta, mass and energies of a particle. (This is as opposed to just automatically using the measurements from the detector.)

- **Calculations of useful variables -** Ntuple production code also performs calculations of many quantities useful for physics analysis. This is helpful not only because it saves the user the trouble of doing it herself, but also because in many cases the information required to do the calculation is not available at the ntuple level - for example, to calculate LAT you need to know which crystals make up a given cluster, and this is listed in the Micro database but may not be in an ntuple.

[10] This is just the standard programming procedure- compiling turns it into a program or *executable* that can actually do something and linking connects the user code to the necessary libraries of pre-existing software.

The ntuple production code then stores the data in ntuples. An ntuple typically contains several thousand events. For each event, there is a list of all the particle candidates in the event, organized into blocks according to their hypothesized identities: a block of B^0 candidates, a block of J/ψ candidates, a block of π^0 candidates, and so on. For each particle candidate, the relevant quantities from the reconstruction are stored, including measured kinematic quantities like momentum and mass, and the ID numbers and (probable) identities of its daughters.

2.8 Independent analysis

It is once the ntuples are produced that the user really takes over the analysis. The ntuple production job creates ntuples as **.hbook** files for analysis in PAW (Physics Analysis Workstation)[14], or as **.root** files for analysis in ROOT[15]. Once the data is in ntuple form, the user has her very own copy of the data, and can use PAW or ROOT software on any computer to do offline analysis.

Individual researchers typically focus on a single decay or other physics process of interest; and a typical analysis involves using tight final selection criteria to isolate and study the decay.

2.9 Simulated data (Monte Carlo)

Simulated data sets, created with "Monte Carlo" (MC) software, are a very important tool in particle physics experiments, providing a way to test whether experimental results are consistent with theoretical predictions. "Monte Carlo" refers to the standard method for producing simulated data in particle physics (and many other disciplines). The Monte Carlo method generates a random set of events distributed according to input probability density functions (PDFs), which reflect our current knowledge of particle physics. The probability density functions are based on world-average values for the properties of the relevant particles (such as masses, lifetimes and branching ratios); and on the laws of particle physics (such as conservation laws and quantum field theory). If the theoretical input into the probability density functions matches reality, then the real and simulated data sets should agree. If they differ, the first thing to check is that there are no mistakes in the simulation. If there are no mistakes, then differences between real and simulated data could indicate that the theory is wrong or incomplete.

Simulations are used to model the signal and background distributions for the decay of interest[11]. A typical approach is to test and optimize an analysis strategy on Monte Carlo events, before looking at real data. This works because the simulated data set

[11] The "signal" is the event you are interested in. "Background" is everything else. In my analysis, the signal is $e^+e^- \to \phi\gamma\,;\varphi \to K^+K^-$.

comes with "truth" information, so that in addition to the reconstruction of an event you also know what really happened. In particular, one can determine whether a given event is really signal or background so you can use Monte Carlo data sets to find ways to select signal events while rejecting background. BaBar uses a GEANT-based [12] simulation with four stages:

1. *Generation of the underlying physics event.* The output of this stage are the four-vectors of the initial, intermediate and final-state particles, with some "smearing" of the beam energies and collision coordinates to make things more realistic.

2. *Simulation of the particle interactions and the detector response.* The output consists of the signals ("*GHits*") that the imaginary particles produce in the imaginary sub-detectors, as well as truth information about the particles that produced the signals. Effects like multiple scattering and energy loss are taken into account.

3. *Overlaying of backgrounds and digitization of the energy deposits.* This stage models the detector response, taking the *GHits* and translating them into imaginary FEE signals. It also mixes in backgrounds measured from real data.

4. *Reconstruction.* Reconstruction is done in essentially the same way as for real data.

The simulations are set up so that real and simulated data look exactly the same as far as the OPR is concerned. The only difference between a real data ntuple and a Monte Carlo ntuple is that a Monte Carlo ntuple has an additional "truth block" which can be used to access the truth information about the particles.

To be confident in the selection criteria that one chooses based on studying the MC events, it is necessary to know how well the simulation imitates the real events. These kinds of studies are usually done using so-called "Control sample" events. In a control-sample event, we know exactly what the final products are, so we can see what really happens in the detector when there is such an event. By comparing what we see in real events with what we get from simulation, we determine how well one can rely on the MC for events containing these final products. We can also extract the correction factors that should be applied to the MC in order to have better agreement with the real data. This thesis contains such a study, using the $e^+e^- \to \varphi\gamma; \varphi \to K^+K^-$ control-sample to study the extra charged tracks in an event, the shape and quantity of extra energy deposited in the EMC and how well the MC simulates these. It also uses this control sample to study the efficiency of the kaon identification system used in BaBar experiment.

3

Measurement of CP violation in BaBar

3.1 The CKM Matrix

There are two necessary conditions that allow a certain level of CP violation in the Standard Model (SM) [16]:

1. *Quark mass eigenstates* ≠ *Quark weak eigenstates*

u	c	t
d	s	b

Table 3-1: Mass eigenstates

u	c	t
d'	s'	b'

Table 3-2: Weak eigenstates

What this means is that the quark generations are "skewed" for purposes of weak interactions. The weak interaction transforms the three down-like quarks from their mass eigenstates (q) to their weak eigenstates (q')[3],[4],[17]. The coupling of these two eigenstates is described by a unitary matrix called the CKM matrix[3] V_{ij}, after the three physicists, Cabibbo, Kobayashi and Maskawa:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

Equation 3-1

2. *There exist at least 3 generations with unequal quark masses*

   The CP violation possibility arises from the presence of a complex phase in the SM Lagrangian. The complex phases appear in the unitary matrix relating the quark mass eigenstates with their weak eigenstates.

   For n generations, the mixing matrix V is an $n \times n$ unitary matrix. Such a matrix can be parameterized in terms of $\frac{n(n-1)}{2}$ rotation angles and $\frac{(n-1)(n-2)}{2}$ physically independent complex phases. Therefore, we need at least three generations of quarks to have a complex phase, allowing CP-violation.

If we apply the unitarity constraint to the first and third columns, we get:

$$V_{ud}V_{ub}^{*} + V_{td}V_{tb}^{*} = -V_{cb}^{*}V_{cd}$$

Equation 3-2

This relation can be represented by a triangle, called the unitarity triangle, in the complex plane, Figure 3-1.

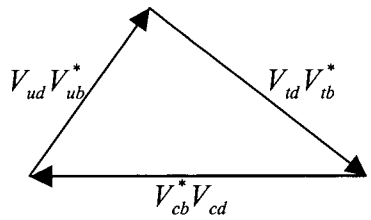

Figure 3-1: The unitarily triangle of the CKM matrix

In the standard parameterization V_{cb} is real and V_{cd} is almost entirely real, then:

$$V_{cb}^{*}V_{cd} = V_{cb}V_{cd} \approx \text{A real number}$$

Equation 3-3

So we could orient the unitarity triangle so that the $V_{cb}^{*}V_{cd}$ side lies on the real axis, as shown in Figure 3-2.

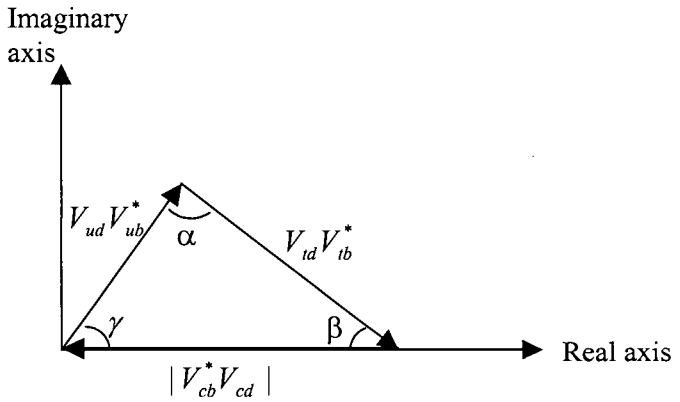

Figure 3-2: α, β and γ positions in unitarity triangle

The sides of the triangle correspond to magnitudes of the CKM matrix elements, and can be measured via the rates of the processes involving these elements. The angles of this triangle, α, β and γ, are related to the CKM phase and can be determined by measuring CP-violating asymmetries. They are given by the following relations:

$$\mathrm{Im}(\frac{V_{ub}^*}{V_{ub}}) = \sin(2\gamma)$$

$$\mathrm{Im}(\frac{V_{td}}{V_{td}^*}) = \sin(2\beta)$$

$$\mathrm{Im}(\frac{V_{ub}V_{td}}{V_{ub}^*V_{td}^*}) = \sin(2\alpha)$$

Equation 3-4

For more information on parameterization of the CKM matrix and the unitarity triangle refer to [17],[18], and [19].

The standard model (SM) predicts CP asymmetries in neutral B decays into certain CP eigenstates. These asymmetries are fully determined by the values of the three unitarity angles; therefore, their measurement will test standard model predictions and probe for physics beyond the standard model.

3.2 CP Violation in the Neutral B System

If we denote the two mass eigenstates of the B system with B_L and B_H and the weak eigenstates with B^0 and $\overline{B}^0$ then we can write [20]:

$$|B_L\rangle = p|B^0\rangle + q|\overline{B}^0\rangle$$
$$|B_H\rangle = p|B^0\rangle - q|\overline{B}^0\rangle$$

Equation 3-5

where p and q are chosen such that B_L and B_H are the eigenvectors of the Schrödinger Hamiltonian:

$$H = \begin{pmatrix} M & M_{12} \\ M_{12}^* & M \end{pmatrix} - i/2 \begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix} = \mathbf{M} - (i/2)\mathbf{\Gamma}$$

Equation 3-6

M is the mass matrix and Γ describes the exponential decay of the system. The constraints on p and q are given by the following eigenvalue equation:

$$\mathbf{M} - \frac{i}{2}\Gamma \begin{pmatrix} p \\ \pm q \end{pmatrix} = (M_{L,H} - \frac{i}{2}\Gamma_{L,H}) \begin{pmatrix} p \\ \pm q \end{pmatrix}$$

Equation 3-7

Due to $B^0 \leftrightarrow \overline{B}^0$ mixing, the initially pure B^0 state produced (for example in the decay of Υ(4S)), develops a $\overline{B}^0$ component as time passes. If we denote the state of B after passing time t with $B^0_{phys}(t)$, then, the time dependent CP asymmetry is defined as:

$$a_{fCP}(t) \equiv \frac{\Gamma(B^0_{phys}(t) \to f_{CP}) - \Gamma(\overline{B}^0_{phys}(t) \to f_{CP})}{\Gamma(B^0_{phys}(t) \to f_{CP}) + \Gamma(\overline{B}^0_{phys}(t) \to f_{CP})}$$

Equation 3-8

From this equation it can be shown:

$$\Gamma(B^0_{phys}(t) \to f) = |\langle f|B^0\rangle|^2 \, e^{-\Gamma t}[1 - \mathrm{Im}(\lambda)\sin(\Delta m t)]$$
$$\Gamma(\overline{B}^0_{phys}(t) \to f) = |\langle f|\overline{B}^0\rangle|^2 \, e^{-\Gamma t}[1 + \mathrm{Im}(\lambda)\sin(\Delta m t)]$$

Equation 3-9

Where Δm is the mass difference between the two mass eigenstates and λ is defined by:

$$\lambda = \frac{q}{p}\frac{\langle f_{CP}|\overline{B}^0\rangle}{\langle f_{CP}|B^0\rangle}$$

Equation 3-10

Substituting Equation 3-9 and Equation 3-10 into Equation 3-8, we obtain:

$$a_{fCP}(t) = \mathrm{Im}(\lambda)\sin(\Delta m t)$$

Equation 3-11

Equation 3-12 demonstrates an example of decays of neutral B mesons to CP eigenstates which could be used to measure α, β and γ [20]:

$$B_d \to \pi^+\pi^- \text{ In which } a_{fCP}(t) \propto \mathrm{Im}(\lambda) \approx \sin(2\alpha)$$
$$B_d \to \psi K_s \text{ In which } a_{fCP}(t) \propto \mathrm{Im}(\lambda) \approx -\sin(2\beta)$$
$$B_s \to \rho K_s \text{ In which } a_{fCP}(t) \propto \mathrm{Im}(\lambda) \approx -\sin(2\gamma)$$

Equation 3-12

Note that the B_s is not produced at the Υ(4S) resonance. (As will be mentioned in the next section, BaBar produces $B\overline{B}$ pairs through the decay of Υ(4S))

One of the main goals of the BaBar experiment is to extract a numerical value for $\sin(2\beta)$, as a probe to measure CP violation in the neutral B decay. Next section provides a discussion on how this goal is pursued in BaBar experiment.

3.3 CP Violation Measurement in BaBar

As we know by now, the crucial measurement of CP violation in BaBar is a comparison of the time dependent rates for B^0 and $\overline{B}^0$ decay to a self-conjugate state, like ψK_s. For maximum cross-section and clean events (i.e. no other tracks in the b events), the best method is to produce $B\overline{B}$ pairs through:

$$e^+e^- \to \Upsilon(4S) \to B\overline{B}$$

Clean experimental results require events in which one B-meson decays to a CP eigenstate that is fully reconstructed and the other B-meson is tagged as a B^0 or a $\overline{B}^0$ by its decay products: a charged lepton, a charged kaon or other flavor sensitive features such as a low momentum pion from a D^* decay.

If we know the flavor of one B at $t = t_1$ (tag B) then we also know the flavor of the other one at that exact time. What we do then is to look for the other B to decay to a self-conjugate state such as ψK_s^0, Figure 3-3. After reconstructing both flavors, knowing the distance between the two vertices and the velocity of the Bs, we can calculate $t = t_2 - t_1$ and use it in Equation 3-11 to find the asymmetry.

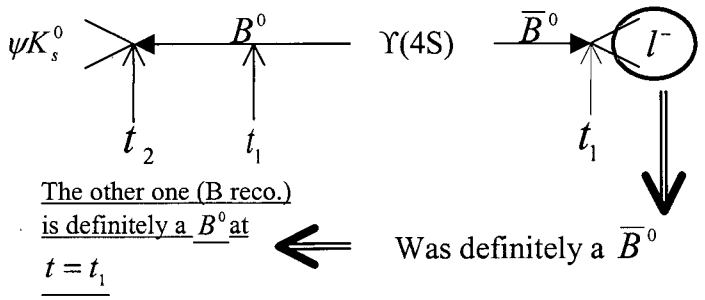

Figure 3-3: Reconstruction of tagged B and B reco. For the tagged B we look for some flavor sensitive products such as a charged lepton. The B reco. is reconstructed using a self conjugate state such as ψK_s^0.

In principle, by measuring the position at which B decays, and knowing its velocity, we can determine the time at which it decayed. But since B has a very short life time ($\tau \approx 1.5 \times 10^{-12}$ sec), and it hardly has a significant momentum in the Υ(4S) rest frame, it would be very hard to measure the time difference between the two B s if the Υ(4S) is not moving in the detector's rest frame.

The solution is provided by using asymmetric beam energies so that Υ(4S) is boosted in the lab frame. This allows the two B s to travel a measurable distance before they decay, Figure 3-4.

In the BaBar experiment, the energy of the electron beam is 9 GeV while that of the positron beam is 3.1 GeV, giving the total Lorentz boost of $\beta\gamma$ =0.56 to Υ(4S).

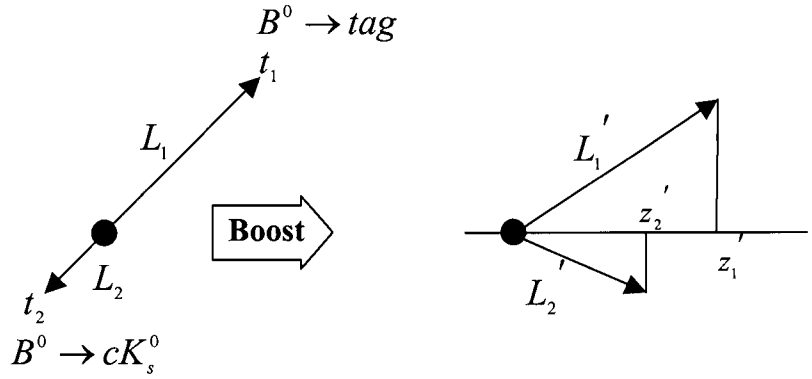

Figure 3-4: Boosting the Υ(4S) rest frame.

Now to calculate the asymmetry $a_{fCP}(t)$ one must measure four distinct configurations depending on whether the tag or the CP eigenstate decays first:

$$n_1 : f_{CP}(t_1) f_B(t_2)$$
$$n_2 : f_B(t_1) f_{CP}(t_2)$$
$$\overline{n}_1 : f_{CP}(t_1) f_{\overline{B}}(t_2)$$
$$\overline{n}_2 : f_{\overline{B}}(t_1) f_{CP}(t_2)$$

Equation 3-13

CP violation produces a $(t_2 - t_1)$ distribution. This distribution would be different for n_1 & $\overline{n}_2$ pair from that of $\overline{n}_1$ & n_2. The asymmetry here is:

$$A_{CP} = \frac{(n_1 - \overline{n}_1) - (n_2 - \overline{n}_2)}{n_1 + \overline{n}_1 + n_2 + \overline{n}_2}$$

Equation 3-14

However, since the $\Upsilon(4S)$ is in a $J = 1$ state and the B mesons are spin-0 particles, conservation of angular momentum forces the two B mesons into a state with orbital angular momentum of 1. ($L = 1$; p-wave). When the $B^0\overline{B}^0$ pair is produced, B^0 and $\overline{B}^0$ are in orthogonal states, so if we see one of them decay as B^0 we know at that instant the other one must have been a $\overline{B}^0$. So the time t in Equation 3-9 should be measured with respect to the time at which one of the neutral B mesons decayed to a "tagging" state. If we assume that the B which decays first, decays to a tagging state and the other one decays to f_{CP}, then $t = t_{CP} - t_{tag} > 0$. On the other hand if the first one to decay is the one that decays to f_{CP}, then $t = t_{CP} - t_{tag} < 0$.

Using Equation 3-9 and Equation 3-12 and substituting $\tau = \frac{1}{\Gamma}$ we can write the decay probability of the B meson as:

$$f(t) = \frac{e^{-t/\tau}}{4\tau}(1 \pm \sin(2\beta)\sin(\Delta m t))$$

Equation 3-15

Where "+" refers to the tag being a B^0 and "-" is to $\overline{B}^0$.
Figure 3-5 shows the plot of the decay probability for both cases.

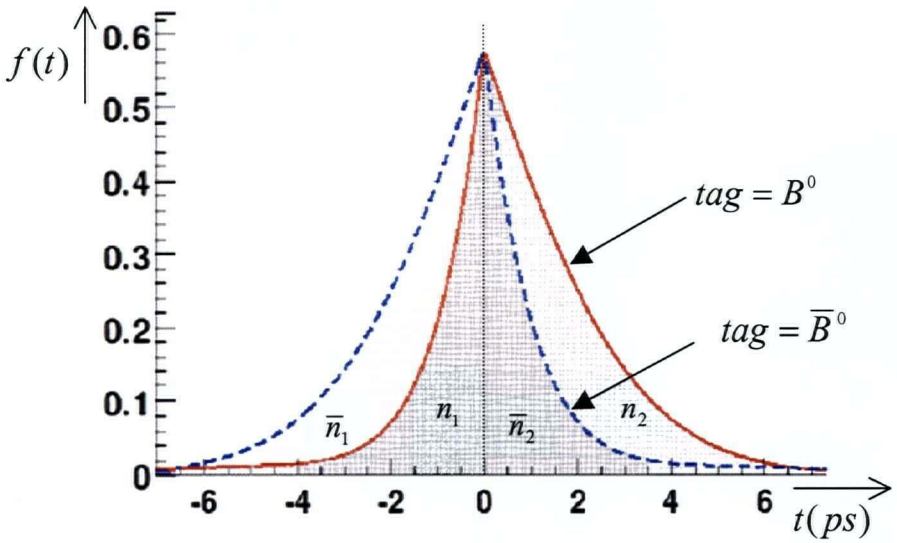

Figure 3-5: This Figure shows the ideal decay probability distribution without taking into account the tagging errors, errors on Δt and errors in selecting B_{reco}.

If we take into account the tagging error, the errors in measuring Δt and the errors in selecting B_{reco} the distribution will look like Figure 3-6:

The extracted asymmetry, as expected, is a sine function. The real experimental result looks like what is presented in Figure 3-7 [21]. The value of $\sin(2\beta)$ is extracted by fitting a function to the resultant asymmetry.

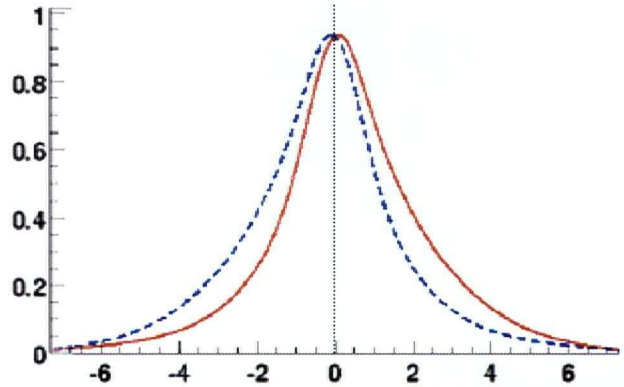

Figure 3-6: This figure shows the plots of decay probability distribution taking into account the various experimental error effects mentioned above.

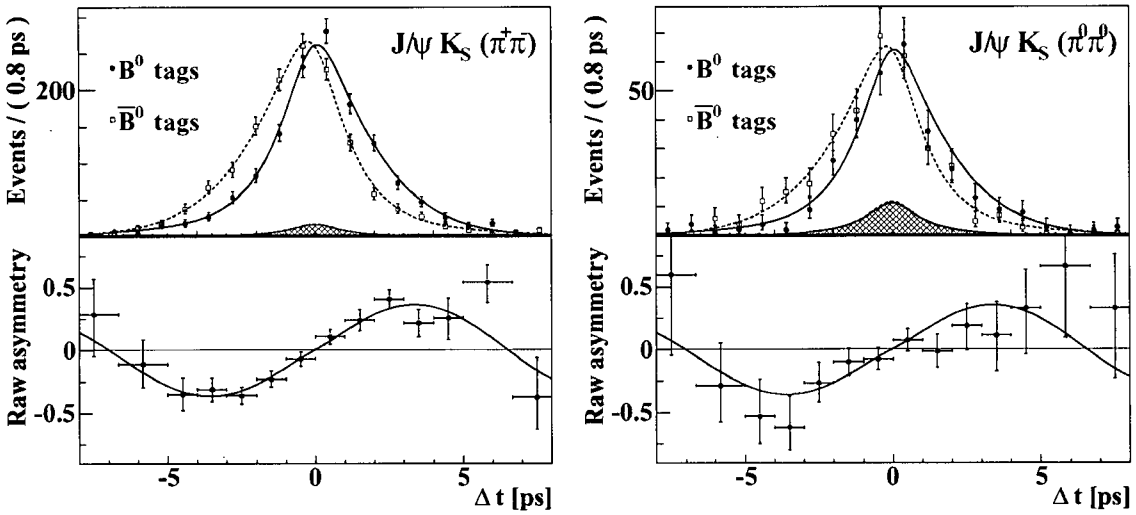

Figure 3-7: Top: The decay probability distribution for when the tag particle is B^0 or $\overline{B}^0$. The reconstruction modes are mentioned on top right of the figure. Bottom: The corresponding extracted asymmetry.

4

Rare B Decays

There are many studies being done in the BaBar experiment looking for rare B decays, since these decays provide a highly sensitive probe for physics beyond the Standard Model (SM). One such study is a search for $B^{\pm} \to K^{\pm} \upsilon \overline{\upsilon}$ using semi-leptonic tags [23] by Paul D. Jackson and Bob Kowalewski of the University of Victoria.

This chapter provides a very quick overview of their analysis (without mentioning the results) and how it led to the study presented in this thesis.

4.1 Search for $B^{\pm} \to K^{\pm} \upsilon \overline{\upsilon}$

FCNC (Flavor Changing Neutral Current) decays are absent at the tree level in the SM, appearing only at the loop level and are therefore highly suppressed. Such transitions provide an opportunity to test the SM and offer a complimentary strategy to search for new physics by probing the indirect effects of new particles and interactions. In the SM, the quark level process $b \to s \upsilon \overline{\upsilon}$ proceeds via penguin or box diagrams, as is shown in Figure 4-1.

The diagrams are similar to the $b \to s l^{+} l^{-}$ decay (aside from the photonic penguin), which has been previously analyzed in BaBar[22]. The undetectable neutrinos complicate things, however, and hence the analysis strategy will differ somewhat. New physics models, which may enhance the rate of $b \to s \upsilon \overline{\upsilon}$, include minimal super symmetry (SUSY), multi-Higgs doublet model (with neutral flavor conservation, these models have extra charged Higgs particles replacing the SM W-bosons in the box diagram), leptoquarks, SUSY models with R parity and topcolor models[24],[25].

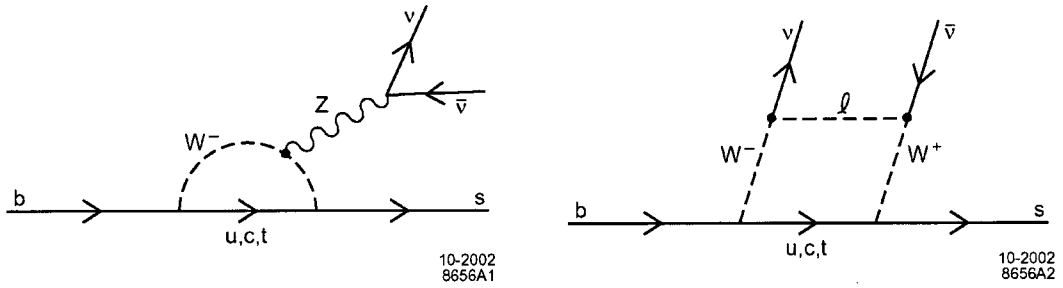

Figure 4-1: Standard Model Feynman diagrams for $b \to s\upsilon\overline{\upsilon}$

An experimental search for the inclusive quark level process $b \to s\upsilon\overline{\upsilon}$ is extremely difficult due to the presence of two unobserved neutrinos. This limits the available kinematic constraints that can be exploited to suppress other B backgrounds. Instead, we search for specific exclusive modes, a method which relies on the reconstruction and removal of the daughters of the other B in the event. The method that has been chosen to carry out this procedure is by tagging the companion B through $B \to Dl\upsilon X$ decays. Such analyses are referred to as "recoil" analyses since one of the B meson decays in the event is reconstructed and then whatever is recoiling against that reconstructed system is studied by searching for consistency with a particular decay of interest. These methods are best suited to the B-factory environment.

4.2 Analysis Overview

The choice of analysis method was guided by several factors. The B reconstruction strategy was strongly motivated by the physics goals of the analysis. Due to the rare nature of the signal, a tagging reconstruction method with both a high efficiency and purity was required.

The decay $B^- \to D^0 l^- \overline{\upsilon}(X)$ is reconstructed in which l denotes either an electron or a muon. After imposing kinematic requirements on the $D^0 - l^-$ combination, the X is usually either nothing or a soft transition pion or photon from a higher mass charm state. The subsequent D meson decay is reconstructed as either $D^0 \to K^-\pi^+$, $D^0 \to K^-\pi^+\pi^+\pi^-$ or $D^0 \to K^-\pi^+\pi^0$. These D^0 decay modes are chosen since they provide both the highest statistics hadronic decay modes and they are the cleanest. Using this technique, neutral modes can be studied via $\overline{B}^0 \to D^+ l^- \overline{\upsilon}$ where $D^+ \to K^-\pi^+\pi^+$. Also, if combining a found D^0 with a low momentum charged pion forms an acceptable $D^{*+} l^- \overline{\upsilon}$ candidate, it is used as a $\overline{B}^0$ tag in place of the B^- tag candidate. Only these reconstructed B^- tags are considered suitable to perform a search for $B^+ \to K^+\upsilon\overline{\upsilon}$.

The search for $B^+ \to K^+\upsilon\overline{\upsilon}$ is performed as a blind analysis by removing all of the events which lie within a "blind region" in specific sensitive selection variables from the analysis chain. Backgrounds in the signal region are estimated from various "sideband" samples prior to unblinding.

4.3 Particle Identification

The hadrons and leptons in the decay products are required to pass different particle identification criteria. In particular, Kaons reconstructed as D meson daughters are required to pass the *KMicroNotPion* selection whereas signal candidate (or primary) Kaons must pass a very tight selection. For the kaon ID the SMS Kaon selector, which is based on the formation of likelihood ratios using quantities from the SVT, DCH, and DIRC, is used.

Motivated by this analysis, and as a part of my research, I have studied the Kaon identification system efficiency, which is described in full detail in Chapter 7.

4.4 Signal selection

Signal candidate events are selected from events in which a suitable tag B candidate is found, by examining the detector activity that is not associated with the daughters of the reconstructed tag. The recoiling "signal candidate" is presumed to be composed of all reconstructed objects (track and clusters) not associated with the tag B. Under the hypothesis that the $B^+ \to K^+ \upsilon\overline{\upsilon}$ decay is the only signal-side[12] activity, and assuming a perfect detector, one would expect only a single track associated with the signal Kaon. In practice, however, additional activity, mostly in the EMC, is present due to both detector and physics effects (such as bremsstrahlung and hadronic split-offs and background from the beam). The basic signal-side selection involves searching for a single track that is consistent with a Kaon, while limiting additional activity within the detector volume. The candidate track must satisfy the following:

- The track must lie within the angular acceptance of the DIRC, $0.469 < \theta < 2.457$
- The candidate track must pass *tight* Kaon identification from the *PIDKaonSMSSelector.*
- The charge of the candidate Kaon must be opposite to that of the lepton used on the tagging side

As the Kaon is the only signal candidate particle directly searched for in this analysis, other activity in the detector due to charged tracks or neutral clusters is considered to be background. There are several quantities which can be used to separate signal from background after the tagging side has been removed from the event, including:

[12] When they make reference to a "side" with regards to the recoiling signal candidate and the tag this is only to distinguish the two hemispheres in which one expects to locate the thrust axes of the B's.

- Number of reconstructed charged tracks: any event with extra charged track(s) is rejected.
- Extra neutral energy in the calorimeter not assigned to the tag B.
- Momentum of the Kaon candidate track.
- Number of remaining K_L^0 candidates reconstructed in the EMC and IFR
- The polar angle distribution of Kaon candidate tracks.
- Various event shape variables.

The second item in the list above will be addressed separately in 4.4.2, where it is explained how the extra neutral energy in the calorimeter is used to select signal candidates and suppress background.

4.4.1 Kaon identification

Kaon selection (and issues relating to it in BaBar) has been outlined in section 4.3 and analyzed in depth in [26].

The *PIDKaonSMSSelector* is based on likelihood ratios mainly between the particle types π, K and p. The tight Kaon selection is optimized to keep the misidentification rate below 2% up to a momentum of 4 GeV/c. Kaon identification is based primarily on Cherenkov radiation information from the DIRC for candidate Kaons with a momentum of p>0.6 GeV/c. Energy loss measurements (dE/dx) from both the SVT and DCH are used for p<0.6 GeV/c. This will not be relevant for the momentum range of either the signal candidate in the Victoria analysis, or in mine. The identification efficiency is measured-using control samples-to be greater than 95% in the range of 1 GeV/c<p*<3 GeV/c.

In general there are 3 different selectors used for Kaon identification:

- ***PIDKaonSMSSelector***
- ***PIDKaonNNSelector***
- ***PIDKaonLHSelector***

Each one has five different modes based on how tight the selection criteria are, and how often they misidentify other particles as Kaons. These selectors and their modes and efficiencies are fully studied as part of this thesis in Chapter 7.

4.4.2 Neutral Energy

Figure 4-2 shows the total remaining neutral energy in the $B^{\pm} \rightarrow K^{\pm}\upsilon\overline{\upsilon}$ event after all tag-side tracks and neutral clusters have been accounted for. This variable will be referred to herein as *Eextra*.

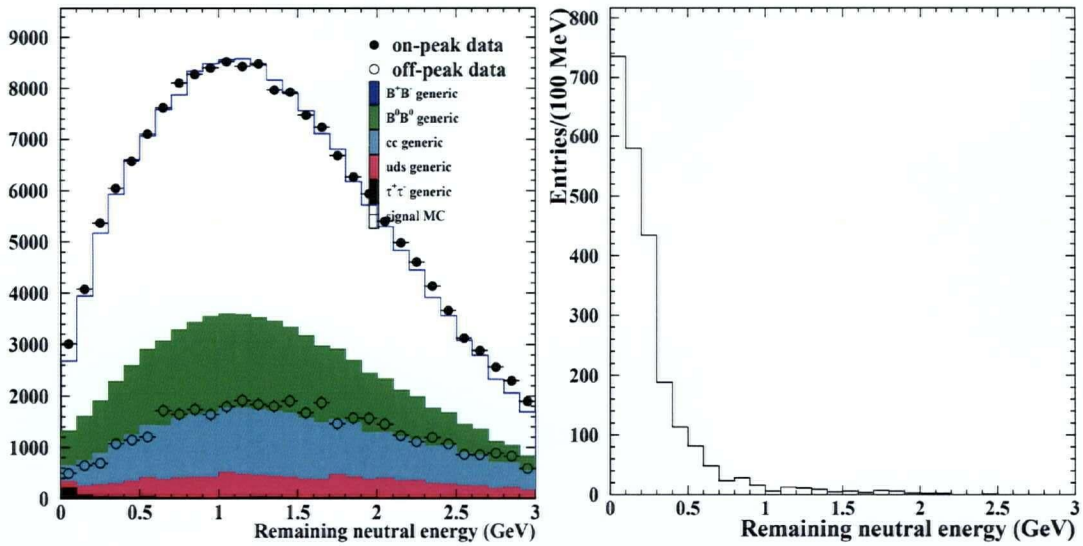

Figure 4-2: A plot of the remaining neutral energy after all the tag side particles have been removed. All D^0 decay modes using all MC ($B\overline{B}, q\overline{q}(q = u,d,s$ or $c)$ and $\tau^+\tau^-$), on-peak and off-peak data. All distributions are scaled to the on-peak data luminosity. The $B^+ \to K^+\upsilon\overline{\upsilon}$ signal MC is plotted on the right.

In a perfect detector the amount of remaining neutral energy for a reconstructed signal event would be zero (modulo any unassociated energy from the tag B). In practice, however, this is not the case because unassigned neutral energy comes not only from unassociated tag-side photons but also from bremsstrahlung photons (from electrons in the tag B), split-offs from hadronic showers[13] (where neutrals can be cast laterally into the calorimeter) and beam related background photons. Other neutral clusters can come from neutral hadrons (K_L^0 's) interacting in the calorimeter. In order to retain a large portion of the signal while minimizing the background, the $B^+ \to K^+\upsilon\overline{\upsilon}$ signal is searched for in the region where *Eextra* is less than 250 MeV. This selection criterion in particular is essential for isolating the signal region from backgrounds and is used in the definition of the "signal box".

The decision to make a cut of 250MeV on *Eextra* is based on the detailed study of signal MC in order to minimize the statistical errors on the signal. One must know precisely how well the MC resembles the real event behavior when one wants to study the *Eextra.* This motivates the other part of my study: to determine how well the MC agrees with real data when studying extra energies in the calorimeter and to extract any correction factors that can be used to improve the agreement.

In addition to the extra neutral energy study, I have also studied the efficiency of cutting on extra charged tracks as a means to extract the signal in the event, since this method is also used in the $B^\pm \to K^\pm\upsilon\overline{\upsilon}$ study, as mentioned in section 4.4.

[13] Details of hadronic showers are not reliably simulated and we need to rely on control samples to understand their behavior.

5

Analysis Overview, Event and Signal Selection

5.1 Analysis Overview

The objective of this analysis is to study detector responses using the $\varphi\gamma$ control sample.

In general, we use control samples to study detector behaviors since they have a clear signal; i.e. we know exactly what the final particles are without requiring them to pass their particle identification algorithms (PID). As an example of such practice and before discussing the $\varphi\gamma$ control sample, the next section illustrates the use of a control sample to study the muon ID system efficiency in BaBar experiment [29]. This will help illustrate how using a control sample provides a good means to study PID system efficiencies and detector behaviors.

5.2 The $\mu^+\mu^-\gamma$ Control Sample

Btamumugamma2Sample is a module in the *BtaPidCalib*[14] package that selects muon tracks for a high purity control sample based on the reaction $e^+e^- \to \mu^+\mu^-\gamma$. This package was created in order to provide a crosscheck on the $e^+e^- \to e^+e^-\mu^+\mu^-$ ($ee\mu\mu$) control sample, which is used to produce the muon ID efficiency tables currently used by BaBar. Analysis of μ ID efficiency differences between the two samples shows a possible contamination in the $ee\mu\mu$ sample at low momentum in the forward end-cap.

[14] *BtaPidCalib* is a package that tags certain events of interest to the particle ID group during OPR processing.

Muons from the reaction $e^+e^- \to \mu^+\mu^-\gamma$, subsequently referred to as $\mu\mu\gamma$, provide a good control sample for measuring muon ID efficiency. The final state γ allows muon tracks to be well distributed over the full solid angle of the detector with a much lower $p-\theta$ correlation than non-radiative di-muons. While the majority of muon tracks have a high momentum, the sample also extends down to below 500 MeV/c.

This sample can be used to calculate the efficiencies of our current muon ID algorithms.

The basic idea is, of course, to find a pure muon sample, so we can then check what fraction of the muons would actually pass the different muon ID selectors, and extract the efficiency of each respectively. The following describes the cuts that are applied to this sample in order to isolate the signal.

The runs used in this analysis were the full Run 1 data set. The muon tracks were selected from two sources in the database. The majority of the events pass the *BGFTwoProng* filter, which is designed to select muon pair events and suppress Bhabhas. Events are also selected from the *DigiL3Prescale*[15] sample with the requirement that there are exactly 2 oppositely charged tracks.

The pair of tracks is then combined with each photon in the event. Photons are taken from the *GoodPhotonLoose* list, and the photon selected is the one which brings the total invariant mass of the 2 tracks + photon closest to the event mass. This is not necessarily the highest energy photon in the event. A cut is applied to the sample requiring the energy of the photon to be at least 1.5 GeV. This eliminates low energy photons, which could be noise. It also significantly reduces the number of events in which the momenta of both tracks are > 4 GeV/c, while preserving events with lower momentum tracks. Since the μ ID tables only extend up to 4 GeV/c, it significantly reduces the amount of unnecessary muon tracks in the sample while still providing the possibility of extending the tables. A further cut is applied, requiring the energy of any other photon in the event to be less than 50 MeV.

A kinematic fit is then applied to the 2 track + photon vertex. The constraints applied to the fit are:

- Geometric constraint with the vertex required to be within the beam spot.
- $\mu\mu\gamma$ invariant mass constrained to the event mass.
- $\mu\mu\gamma$ momentum constrained to the total beam momentum.

The χ^2 probability of the fit is required to be > 0.01. This eliminates virtually all the τ events because of the missing momentum carried by the neutrino in the τ decay. A further cut is applied on the event requiring the calorimeter centroid of the photon candidate to be at least 30 cm apart from each muon centroid to avoid overlap.

A track from an event passing these cuts is added to the list of unbiased muons if the oppositely charged track passes the standard BaBar loose muon selector [30]. No further

[15] One event in every 100 which enters the L3 trigger is allowed to pass without examining the physics criteria of that event. These are mainly used for diagnostic purposes in the trigger system. These events are stored in the central database in the *DigiL3Prescale* list.

cuts are applied to the tracks on this list. They are unbiased with respect to μ ID algorithms and are used to calculate the efficiencies of these algorithms.

After applying these criteria, the purity of the sample was examined. The two major sources of contamination are other radiative di-leptons. However, contamination by $ee\gamma$ and $\tau\tau\gamma$ events have been cut to a very low level, resulting in a sample of high purity. Samples of $ee\gamma$, $\mu\mu\gamma$ and $\tau\tau(\gamma)$ events from the SP3 Monte Carlo were studied to determine the purity of the sample. The results show that the contamination is less than 0.5%. Possible hadronic interactions have cross sections small enough to make contamination insignificant.

While the efficiency of selecting the unbiased muons from the $\mu\mu\gamma$ Monte Carlo is only 1.6%, many of the events do not produce tracks that can help to generate the efficiency tables. However, they have not been cut from the sample to keep the possibility of extending the tables open. Since 97% of the events have both tracks with momentum > 4 GeV/c, this low efficiency does not represent a significant loss of tracks which are of interest in generating the efficiency tables.

After acquiring the sample, we can study the μ ID system efficiency and also check the agreement between the Data and MC to see how well the simulated data resembles the real data.

Briefly, the new $\mu\mu\gamma$ control sample shows very good purity compared to $ee\gamma$ and $\tau\tau(\gamma)$ MC samples. However, the data is in potential low p disagreement with the $\mu\mu\gamma$ MC in both p and θ spectra. When comparing the efficiencies of this control sample with the $ee\mu\mu$ efficiencies from the μ ID efficiency tables, one sees significant discrepancies at low momentum and forward angles for all selectors [29]. Examination of the efficiencies for both samples indicates that the $ee\mu\mu$ sample could have some contamination in this region of phase space; however, due to the discrepancy between the data and MC spectra, this is not a strong conclusion. A new module has been made in *BetaPidCalib* called Btamumugamma2Sample. The module will generate this control sample so that it can be studied further. The detailed results of this study are described in [29].

5.3 $\varphi\gamma$ Control Sample

This study makes use of the $\varphi\gamma$ control sample to study detector behaviors, especially EMC behavior near a charged track, and also to study the kaon identification system efficiency. In the $\varphi\gamma$ control sample, we search for $e^+e^- \to \varphi\gamma$ in which φ decays to two oppositely charged kaons:

$$e^+e^- \to \varphi\gamma$$
$$\hookrightarrow K^+K^-$$

The Feynman diagram of the φ decay is shown in Figure 5-1.

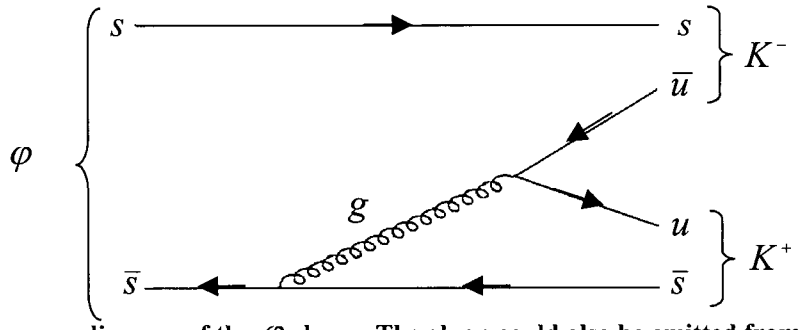

Figure 5-1: Feynman diagram of the φ decay. The gluon could also be emitted from the s quark.

The final particles in this event that we search for, are then K^+, K^- and γ. In an ideal situation these three particles are the only ones that should be present in the detector. Thus, the vector sum of their 4-Momentum should be equal to the 4-Momentum of the center of mass frame, i.e. we shouldn't have any missing energy.

The cuts that are used for the first part of the study- in which the topology and magnitude of the extra energies in the event are investigated- are described in section 5.5.1. As we will see, for this part a tight KID requirement is applied to both charged tracks to assure a clean signal. This does not impose any bias on the study since no PID system is studied at this stage. A comparison is then made to see how well this behavior is simulated in the simulated data. Furthermore the correction factors that could be applied to the MC quantities to improve the agreement between the two are extracted.

We will also look at what percentage of the events *do* have extra charged tracks in them. This would particularly help to clarify whether cutting on extra charged tracks in an event is an efficient way to isolate the signal, as is done in the $B^{\pm} \to K^{\pm}\upsilon\overline{\upsilon}$ study (section 4.4)

We then apply the cuts mentioned in BaBar Analysis Document #633[31] – which gives a recipe of how to choose "good" clusters and "good" tracks - and compare the Data/MC agreement again to see if we could achieve a better agreement between data and MC after applying these cuts to select EMC clusters and charged tracks. The cuts suggested in this document are supposed to improve the data-MC agreement. The idea is that not all clusters/tracks are "good" clusters/tracks. For example some of the clusters could be from different sources of beam background, or in the case of tracks, we might have some "ghost" tracks. Ghost tracks are not real tracks, but somehow the code reconstructs them as tracks. These are fully described in the next chapter.

In the second part of the study, as was mentioned in Chapter 4, the Kaon ID system efficiency in the data and MC is extracted. For this part, only one of the kaons is required to pass through a tight KID while there is no constraint applied to the PID of the other. After isolating the signal and getting a nice clean φ peak signal - which ensures that the second track is also a kaon - we will check what percentage of the second tracks in the whole data set would actually pass the different modes of different kaon selectors. These modes and selectors are described in section 7.5. We will then calculate the corresponding efficiency for each selector using this information.

5.4 Event Selection

This analysis uses the data acquired by the BaBar detector from 1999 to 2003. The data consists of two data sets:

- The "summer 2002" data set includes data collected between BaBar's startup in 1999 and the summer shutdown in 2002. It includes both Run 1 and Run 2 data, and was reconstructed using the 12-series release of the BaBar software. It contains 81.9fb^{-1} of on-peak data, as well as 9.6 fb^{-1} of off-peak data collected at about 40 MeV below the Υ(4S) resonance ($\sqrt{s}$ =10.58 GeV) for background studies.
- Run 3 data set contains data collected between December and June 2003. It was reconstructed using the 12-series release of the BaBar software. It contains 31.2 fb^{-1} of on-peak and 2.4 fb^{-1} of off-peak data.

This analysis does not require a Υ(4S) resonance, so both off-peak and on-peak data[16] collected during Run 1, Run 2 and Run 3 are used. During the ntuple production however, it was apparent that some of the files in the data sets were corrupted, so the integrated luminosity that is used to make the ntuples for this study is slightly less than what one would get if she adds up the luminosities mentioned above. For this thesis, 112.5 fb^{-1} of on-peak and 11.8 fb^{-1} of off-peak data is used, corresponding to an integrated luminosity of 124.3 fb^{-1} in total.

The analysis also uses simulated or Monte Carlo (MC) data to check the agreement between the data and MC and extract the correction factors that could be applied to the MC quantities, to improve the agreement. The MC version used in this analysis is SP5. It was produced during Run 3 set using the release 12 software. (BaBar software nomenclature is explained in Chapter 2)

The data and MC samples are produced in the form of ntuples, the convenient and portable data storage format described in section 2.7. An ntuple serves as a user's private copy of the data. The MC set used in this analysis is:

$$e^+e^- \to K^+K^-\gamma$$

There is no constraint on the invariant masses of kaons in the generator that made this data set, forcing them to be coming from a φ. This means that although most of the events are $e^+e^- \to \varphi\gamma$ with $\varphi \to K^+K^-$, the generator also produces non-resonant $e^+e^- \to K^+K^-\gamma$ events. Therefore, the same set of cuts as was used for the data is applied to this data sample as well. This will ensure the presence of a φ as a mother particle of the two charged kaons. In producing the ntuple, the only criteria used to

[16] On-peak data are collected when the beam energies are tuned on $\Upsilon(4S)$ mass. Off-peak data are collected when the beam energies are 40 MeV below the $\Upsilon(4S)$. The purpose of recording off-peak data is to study the background from non-$B\overline{B}$ events.

select the event is requiring it to pass the *BGFPhiGamma* filter. The cuts that this filter applies are:

- *2 ≤ # Charged Tracks ≤ 5*
- *1 Isolated Photon:*
  - *4 GeV <* E_γ *< 6 GeV*
  - *All tracks satisfy:* $\cos(\phi_{trk} - \phi_\gamma)$ *< 0.8:* to ensure the photon is isolated.
- *2 tracks in opposite hemisphere with invariant mass consistent with* $K_S^0 \to \pi^+\pi^-$ *or* $\varphi \to K^+K^-$*:* The assumption here is that the φ decays to either a K^+K^- pair, or a $K_S^0 K_L^0$ pair. If the second occurs, then the K_S^0 particle itself decays into other particles, mostly a pair of oppositely charged pions, which is what this filter looks for.

It should be remembered, however, that these flags are set after only a *part* of the reconstruction code has been run on the event. So they judge incomplete information, to prevent unnecessary calculations on the events that are not interesting. In other words, they are designed to prevent overloading the PR (prompt reconstruction system). So in this particular case, for example, lots of events could pass the *BGFPhiGamma* flag, without really being $\varphi\gamma$ events. In my ntuple containing 5981588 events, approximately 0.5% of the events are real $\varphi\gamma, \varphi \to K^+K^-$ events. On the other hand, in a sample of real $\varphi\gamma, \varphi \to K^+K^-$ events, approximately 89% of them will pass the flag.

5.5 Signal Selection

This section will describe how the signal for this study is extracted. As mentioned earlier (Section 5.3) there is a small difference in how the signal is extracted between the two different parts of the study, which is explained in detail in this section.

5.5.1 Signal selection for the study of extra energy in the event

The following cuts have been applied to isolate the signal:

- Having two or more charged tracks in the event. All the events already have this condition because they have passed the *BGFPhiGammaFilter*. Figure 5-2.
- Having at least one high-energy photon in the event. The energy of the photon is required to be between 4 GeV and 6 GeV in the center of mass frame (CMF). (Figure 5-2). If two such photons exist in an event, the one with higher energy is

chosen.

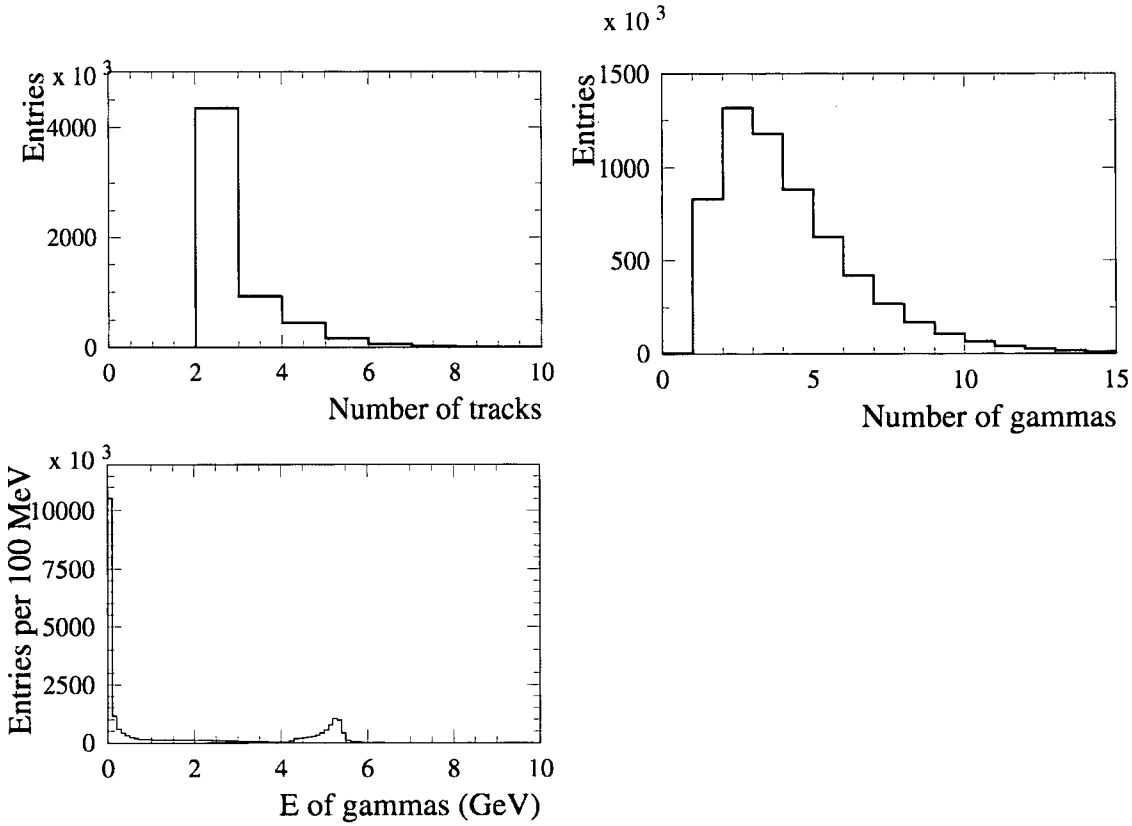

Figure 5-2: Top left: Histogram of the number of charged tracks in the events passing *BGFPhiGamma* flag. Top right: Number of Gammas in the event. Bottom: CMF energy of the gammas in the event.

After the photon is selected, we search for our kaons. We look at all the charged tracks in the event and select a pair of oppositely charged ones that satisfy the requirements described below. The tracks are from the *Charged Track* List. This list does not apply any filtering on the tracks: all reconstructed charged tracks in the event are in this list.

The cuts applied to select the kaons are:

- Candidate tracks should pass *tight* kaon identification from the *PIDKaonSMSSelector* . (The selectors are described in section 7.5)
- Candidate tracks should be roughly back-to-back with the photon: $Cos(Trk,\gamma) \leq 0.5$ (in other words: 0 rad $\leq (Trk,\gamma) \leq 1.047$ rad) (Figure 5-3).

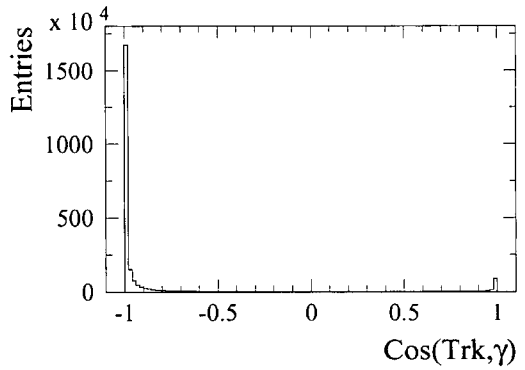

Figure 5-3: The Cosine of the angle between Kaon candidate track and the high energy Gamma.

- The candidate track must be in the angular acceptance of the DIRC ($0.445 \leq \theta \leq 2.656$), and the CMF momentum of the track is required to be greater than 1 GeV/c to ensure that the track makes it to the DIRC. As you can see in Figure 5-4 and in Figure 5-5, in addition to the φ peak, there is a cluster of events at low momentum and low angle that form a mass peak around roughly two times the mass of the kaon. These clusters are the tracks that are misidentified as kaons, since they haven't made it to the DIRC to be identified more precisely. They are actually low momentum pions to which we have assigned the kaon mass. As you can see in Figure 5-6 most of these low momentum tracks are correlated with the low angle ones. By requiring the track to make it to the DIRC, this cluster gets cut off from the signal events.

If more than 1 pair of kaons satisfying all requirements are found, we pick the pair that has the closest invariant mass to the mass of φ. It was found that only about 0.2% of events have more than 1 pair of kaons satisfying the mentioned requirements.

After applying all these cuts, the φ peak made with the two candidate kaons is shown in Figure 5-7.

We can see the best Data-MC agreement is in the region where $m_{\varphi} \leq 1.04$ GeV/c^2. So the last cut used to extract the signal is:

- $m_{\varphi} \leq 1.04$ GeV/c^2.

This cut also ensures that we don't have any $e^+e^- \rightarrow K^+K^-\gamma X$ events in the sample. In this sample, 37860 events pass the cuts mentioned above.

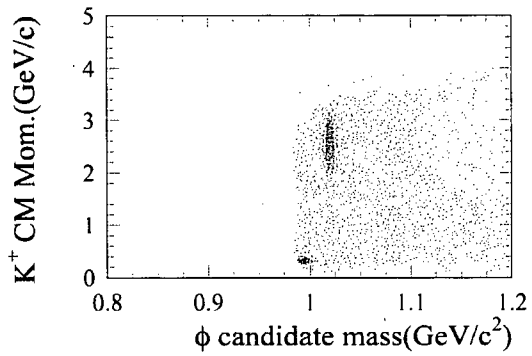

Figure 5-4: As you can see, there is a cluster of events other than the φ peak which have quite a low momentum. That's why the mass they produce is about two times the mass of a kaon.

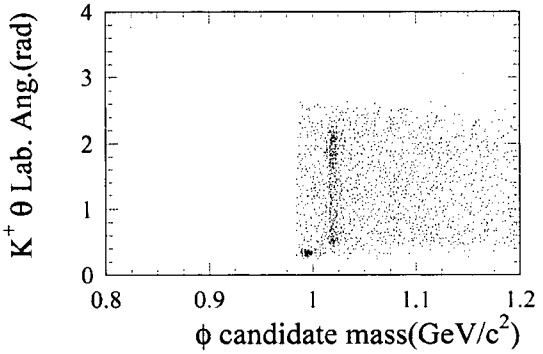

Figure 5-5: The same cluster is seen in this figure for tracks at low angle. These tracks do not actually make it to the DIRC.

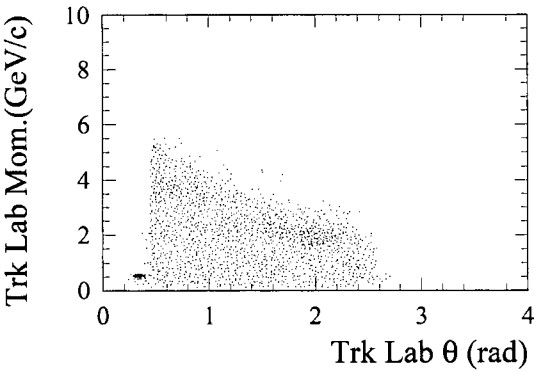

Figure 5-6: The correlation between low momentum and low angle events is apparent in the cluster at the bottom-left corner of the figure.

5.5.2 Signal selection for study of KID system efficiency

The difference between signal extractions in studying KID system efficiency and studying extra energy in the event is that for one we only require *one* of the kaons to pass through the tight *PIDKaonSMSSelector* and to make it to the DIRC. There is no constraint on the PID of the other kaon, neither have we required it to possess momentum and angle requirements to make it to the DIRC. Another difference is that

we don't apply the cut on mass of the φ ($m_{\varphi} \leq 1.04$ GeV/c^2) because we do a fit on the φ mass and we don't want to truncate the tail.

Total number of events that pass the cuts in this part of the study is 65317 from which 32932 events are real $\varphi\gamma, \varphi \rightarrow K^+K^-$ events (about 50%).

5.6 Summary

In this chapter, we discussed why we use control samples, and what our objective are in studying the $\varphi\gamma$ control sample. Later, we discussed the selection criteria to choose an event to be stored in the ntuple, and then how we choose our signal events from the events stored in the ntuple. The next chapters will be a fully detailed discussion of the two analyses (extra energy study and KID system study) and their results.

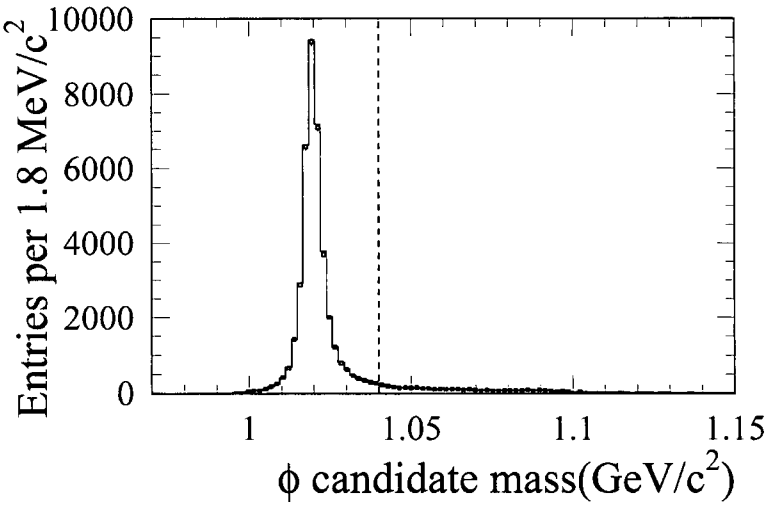

Figure 5-7: φ peak mass distribution after all cuts are applied. The continuous line represents real data while the squares represent the MC. The MC is normalized to the data.

6

Extra Energy Study

6.1 Extra Neutral Energy

In this analysis, "Extra neutral energy" ($E_{Extra-neutral}$) refers to the total remaining neutral energy in the event after all signal tracks and neutral clusters are accounted for. In a perfect detector, this extra neutral energy would be zero[17]. In practice, however, this is not the case since unassigned neutral energy can come from bremsstrahlung photons, Hadronic split-offs (from hadronic showers) or beam related background photons.

It is a common practice in some studies to make a cut on extra neutral energy to isolate the signal. The decision on the amount of energy cut, as mentioned in section 4.4, usually comes from studying the signal MC of the event of interest. In this research, we studied the agreement between data and MC to see how much we could rely on the simulation when extra neutral energy is present. Furthermore, we extract the possible correction factors that can be applied to MC quantities to improve this agreement.

The disagreement is assumed to be largely due to the fact that hadronic and electromagnetic showers near the high energy photons are poorly simulated in BaBar Monte Carlo simulations.

Figure 6-1 and Table 6-1 present Data/MC comparisons for the extra neutral energy in this analysis. As can be seen, the MC underestimates the $E_{Extra-neutral}$ by a factor of 17%. They also provide some information about the percentage of event with almost zero $E_{Extra-neutral}$ ($E_{Extra-neutral} \leq 20\text{MeV}$). Data and MC seem to have a rather good agreement in this region.

All the numerical results are summarized at the end of this chapter.

[17] In a typical analysis we don't really know all the particles in the final state. This sample is unique because we DO know all the final particles, so we expect the extra energy to be zero.

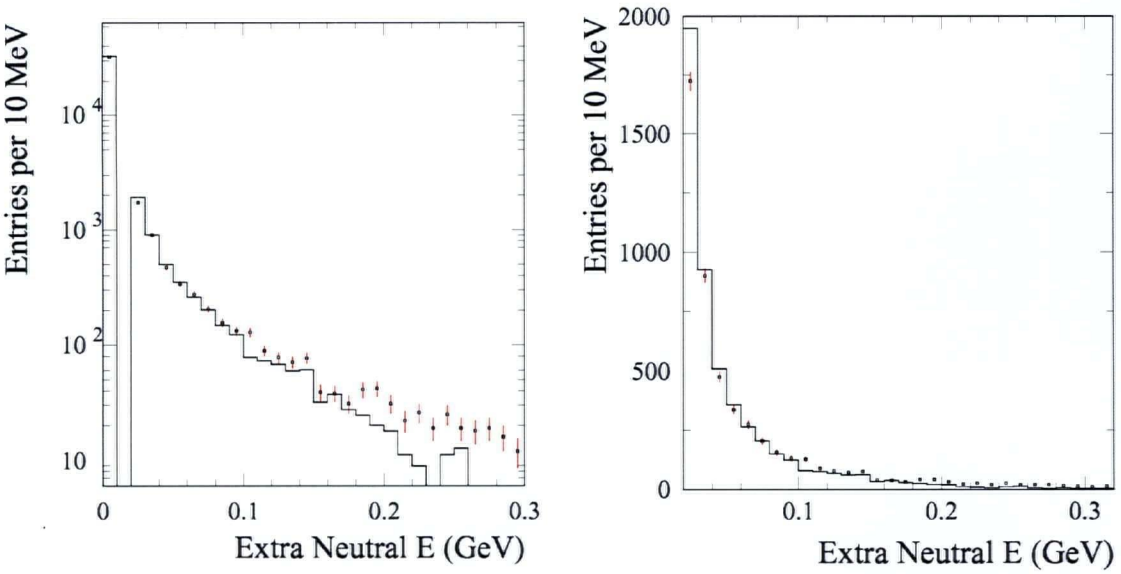

Figure 6-1: Extra neutral energy in the event. The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale. (b) Linear scale for $E_{Extra-neutral} > 20$MeV.

	MC	Data	Correction factor
Average Extra Neutral Energy (MeV)	7.0	8.3	1.2
Fraction of Events with Extra Neutral Energy ≤20MeV	86.9%	85.2%	1.0

Table 6-1: Data/MC comparison of average extra neutral energy in the events. This information indicates that if we want to get the right amount of extra neutral energy in data based on information we obtain by studying signal MC, we have to multiply the MC results by a factor of 1.2.

6.1.1 Neutral clusters near a charged track

Another interesting subject that we will be looking at in this study is the cluster shape and amount of energy deposited in the EMC near the charged kaons. The list that is used in this study to choose the EMC clusters from is *CalorClusterNeutral* List, which is a list of *BtaMicroCandidates* that are multi-bump neutral clusters or single bumps which are not part of a cluster that is matched with a track. So in this study, we only see the energy deposited in the Calorimeter by neutral particles like pions or photons but not charged ones like K^+ s and K^- s.

When a charged track hits the EMC, some energy will be deposited in the EMC due to the interaction between the charged particle and the EMC material components.

If the track is a hadron, the interaction results in production of π^0 s and $\pi^{\pm}$ s. In the case of π^0 s, they would decay into photons, and deposit energy in the calorimeter. Charged pions on the other hand can undergo hadronic interactions and produce hadronic showers[18]. If the tracks were e^+ s and e^- s, they would produce photons directly and deposit energy in the EMC. There is no direct way of knowing exactly if an energy cluster is coming from a charged track interaction at the EMC or not. Usually, a cut is made on the angle between the track and the EMC cluster. Anything inside this defined cone is considered to be coming from the track and so the energy will be associated with the track. Anything outside is considered to be from another source (Figure 6-2).

Conceptually speaking, the problem that could occur is that hadronic showers, like photons from π^0 decay could penetrate deep into the calorimeter before they interact and deposit energy in the EMC. When this happens a cluster is formed outside the defined volume and the energy will not be assigned to the charged track[19]; it will be added to the left over energy in the event. This analysis will provide a study of how often this happens and if it implies any significant effect in measuring the left-over energy in the event. This will be done by looking at the distribution of energy deposited at the EMC near the charged track, after all clusters assigned to the charged track are removed.

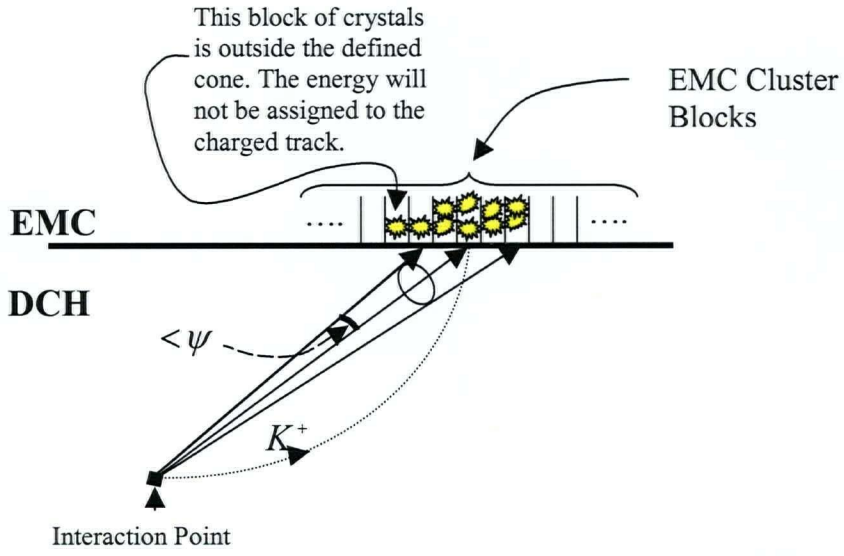

Figure 6-2: Conceptually we can think that all hits in a block of crystals inside the cone defined by angle ψ, is considered to be coming from the charged track. If the cluster lies outside this region, the energy will not be assigned to the charged track[19].

It should be mentioned that from now on in the figures, ψ_{trk} is the minimum of ψ_{K^+} and ψ_{K^-}. As you can see in Figure 6-3, there is a peak near the charged track that indicates some of the clusters outside the defined cone should still be associated with

[18] In a hadron cascade roughly 30% of the incident hadron energy is lost by the break up of the nuclei, nuclear excitation, and evaporation neutrons (and protons), and does not give an observable signal[32].
[19] In reality, however, this might not be quite true as continuous blocks of clusters are considered as one.

the charged track. Otherwise, we would see a constant background due to other activities in the detector that are not spatially related to the charged track. We can also see that the MC underestimates the amount of these hadronic showers. There also seems to be another peak in the MC near the signal gamma, which does not exist in the real data. Most probably, these clusters are from electromagnetic showers of the high-energy photon. This effect seems to be overestimated in the MC, and hence have formed a peak shaped distribution near the signal photon.

By looking at Figure 6-4, we can see that not only the number but also the energy of the hadronic shower clusters are underestimated in the simulated data since there exist many more high energy clusters near the charged track than in the simulated data.

The same conclusion can be derived by comparing Figure 6-5 and Figure 6-6: although the total number of entries in the MC is somewhat higher than the in the real data, and we have roughly the same amount of entries near the charged track in MC as we have in data, when we weight the entries by the amount of energy released in each cluster the MC curves shrinks much more than the data does. This is another sign of the MC underestimating the amount of energy deposited in each cluster.

One other difference that could be mentioned is that in data (Figure 6-5), the curve away from the charged track shrinks much more than the ones near it when we weight the entries by the energy deposited in the clusters. This is due to the fact that the clusters away from the track have far less energy than the ones near it. We do not see the same behavior in MC as we look at Figure 6-6: it is because in the MC, we do not see that big a difference between the amounts of energy deposited near the charged track (Figure 6-4), and the ones away from it.

The summary of some numerical comparisons between data and MC is provided in Table 6-2. In the first part of the table, the events with $\psi_{trk} < 1$ are used to exclude the extra energy due to the gamma. You can see that the MC underestimates the hadronic showers both in number of clusters (by a factor of 23%) and in the amount of energy deposited in those clusters (by a factor of 74%). In the second part, a comparison is made between the energy deposited in the calorimeter near the gamma in data and MC, which shows that the average energy deposited near the gamma is underestimated in MC by a factor of 29%. Also the table shows that the average energy per cluster in MC is less than that of data by a factor of 33%. This explains the apparent contradiction between Figure 6-3 and Table 6-2.

As we can see, in Figure 6-3, there more clusters near the gamma in MC than there are in Data, but Table 6-2 suggests that the energy deposited near the gamma is more in Data than in MC. This apparent contradiction is because although we have more clusters near the gamma in MC, each of them on average has less energy than the clusters in Data.

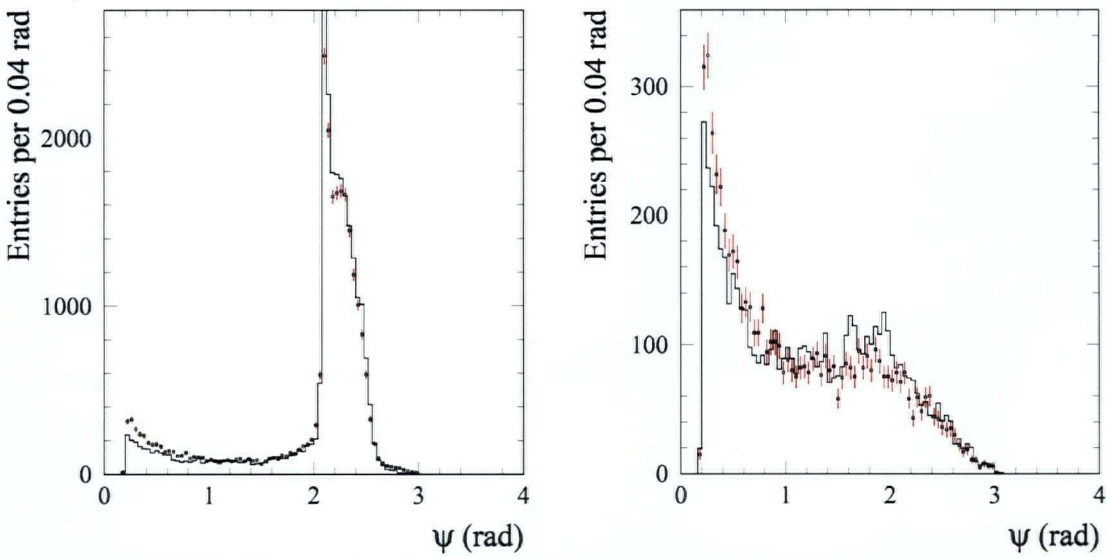

Figure 6-3: The dotted points are Data and the histogram is MC that has been normalized to Data luminosity. (a) ψ_{trk} distribution for all clusters. (b) ψ_{trk} distribution for all clusters except high-energy signal photon.

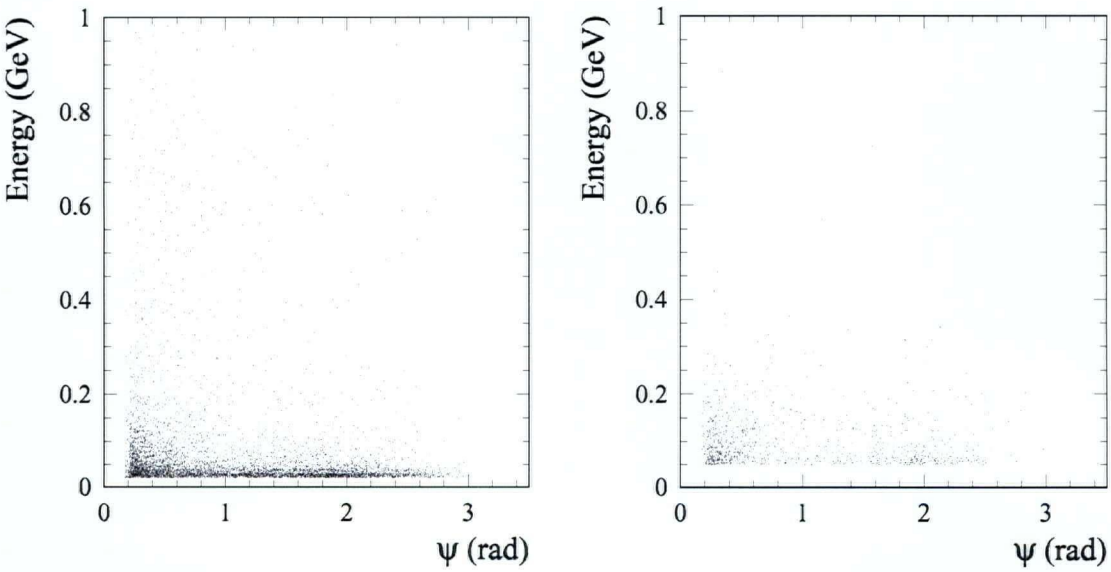

Figure 6-4: (a) The energy released at each cluster versus ψ_{trk} of the cluster for the Data. (b) Energy released at each cluster versus ψ_{trk} of the cluster for the MC.

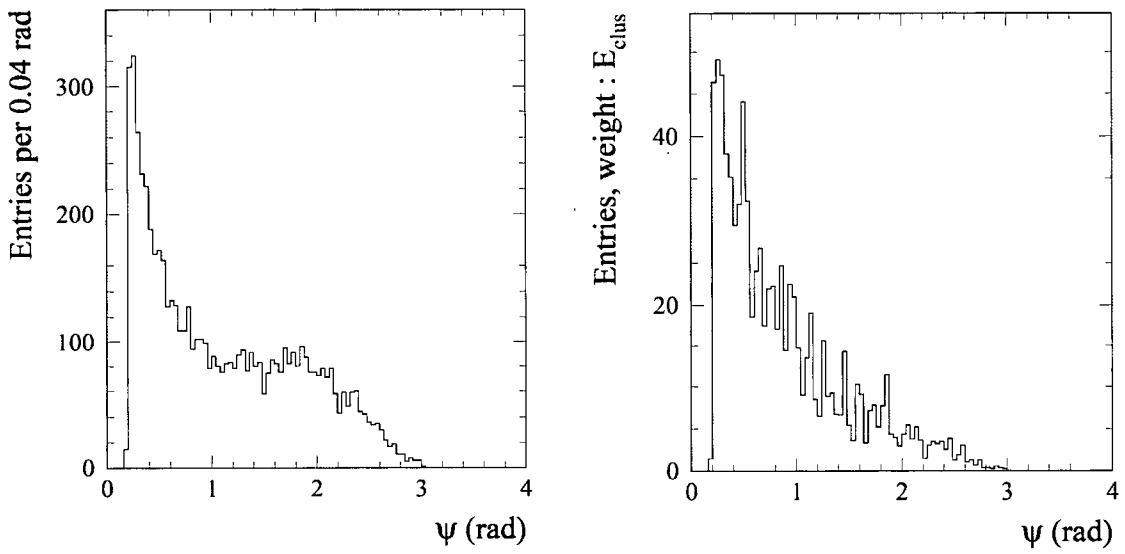

Figure 6-5: (a) ψ_{trk} distribution in data. (b) ψ_{trk} distribution for data when each entry is weighted by the energy of the associated cluster.

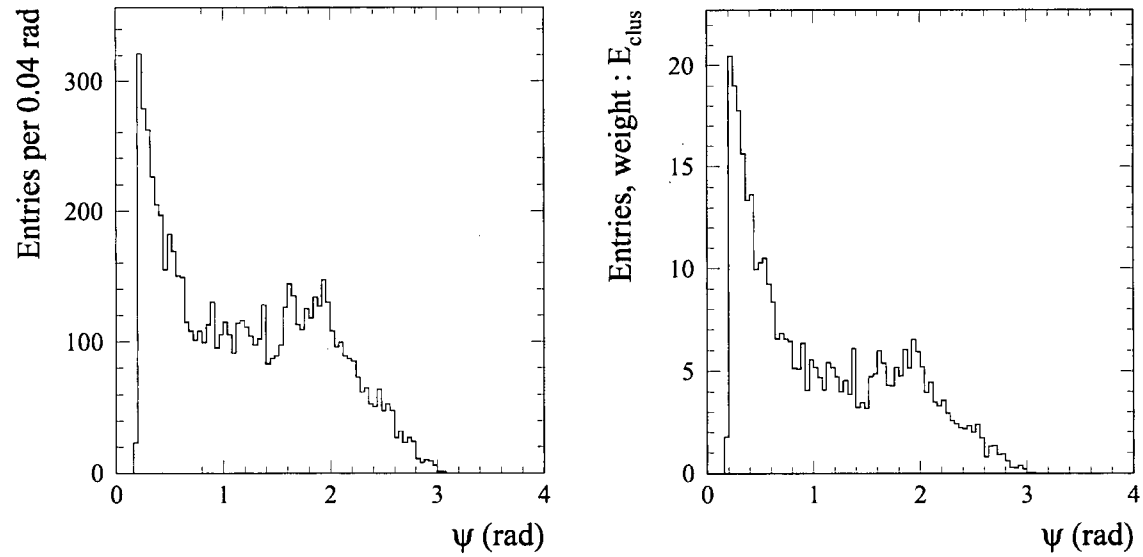

Figure 6-6: (a) ψ_{trk} distribution in MC. (b) ψ_{trk} distribution in MC for when each entry is weighted by the energy of the associated cluster.

		MC	Data	Correction Factor
For $\psi_{trk} < 1$ Rad.	# of clusters/charged track	0.065	0.086	1.3
	Energy deposited (MeV)/charged track	4.0	15.0	3.8
For $\psi_{gamma} < 1$ [20] Rad.	Energy deposited(MeV)/gamma	1.9	2.7	1.4
	Energy deposited(MeV)/cluster	41.1	62.4	1.5

Table 6-2: Data/MC comparison for the number of clusters near a charged track and the high energy photon and the amount of energy deposited near them.

6.2 Extra Charged Tracks and Energy

6.2.1 Extra Charged Tracks

Extra charged tracks are all non-signal tracks in the event. These can come from beam backgrounds or from hadronic or electromagnetic showers in the calorimeter. Some analysts make a cut on extra charged tracks in the event to purify their signal. Figure 6-7 and Table 6-3 demonstrate this variable in our analysis. As you can see, most of the events have zero extra charged tracks. The MC underestimates the number of extra charged tracks, in comparison to the data, but there is a relatively good agreement between the two.

6.2.2 Extra Charged Energy

In this study, extra charged energy is defined as the sum of the energy of the charged tracks in the event except the signal tracks, i.e. the K^+ and the K^-. The energy is calculated using Equation 6-1 which describes the Energy-Momentum relation for any particle with mass *m*. *E* is measured in eV, *m* is measured in eV/c^2 and p is in eV/c with c=1.

A pion mass hypothesis is used to perform the calculation since pions are the most probable charged particles in the detector. The results for data and MC are shown in

[20] ψ_{gamma} is defined exactly the same as ψ_{trk} in Figure 6-2, except that instead of having a charged track, we have a gamma track. So it is defined as the angle between the gamma track and a cluster in EMC.

Figure 6-8 and Table 6-4. As you can see, the MC underestimates the amount of extra charged energy by 29%.

.

$$E^2 = m^2 + \left|\vec{p}\right|^2$$

Equation 6-1

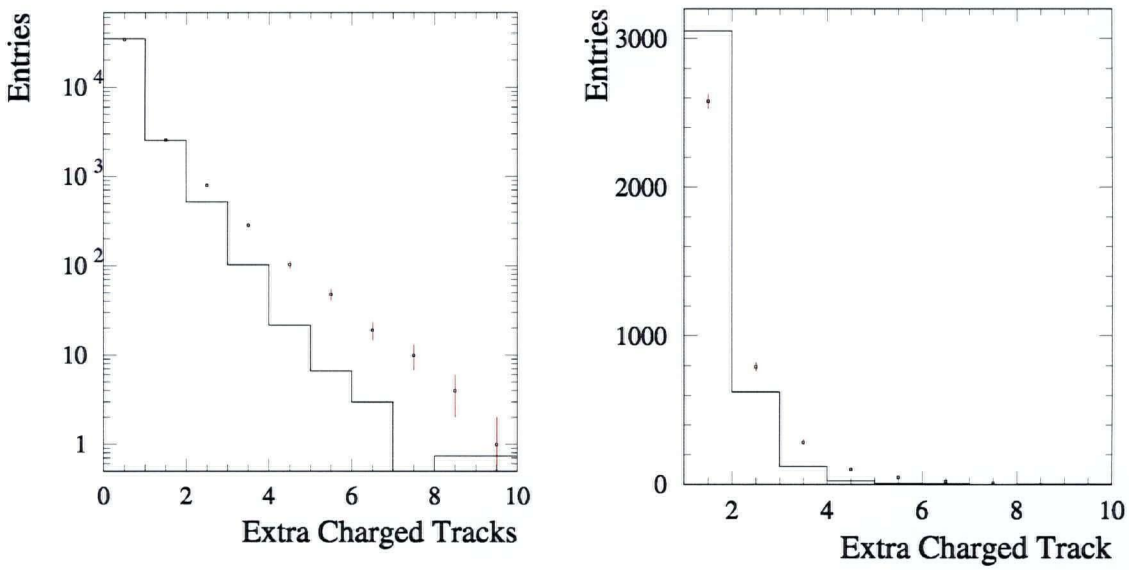

Figure 6-7: The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. Number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale of number of extra charged tracks in the event. (b) Number of extra charged tracks in the event for $N_{Trk-extra} \geq 1$.

	MC	Data	Correction Factor
Percent of events with extra charged tracks	8.3%	10.1%	1.2

Table 6-3: A comparison between number of extra charged tracks for events in data and MC.

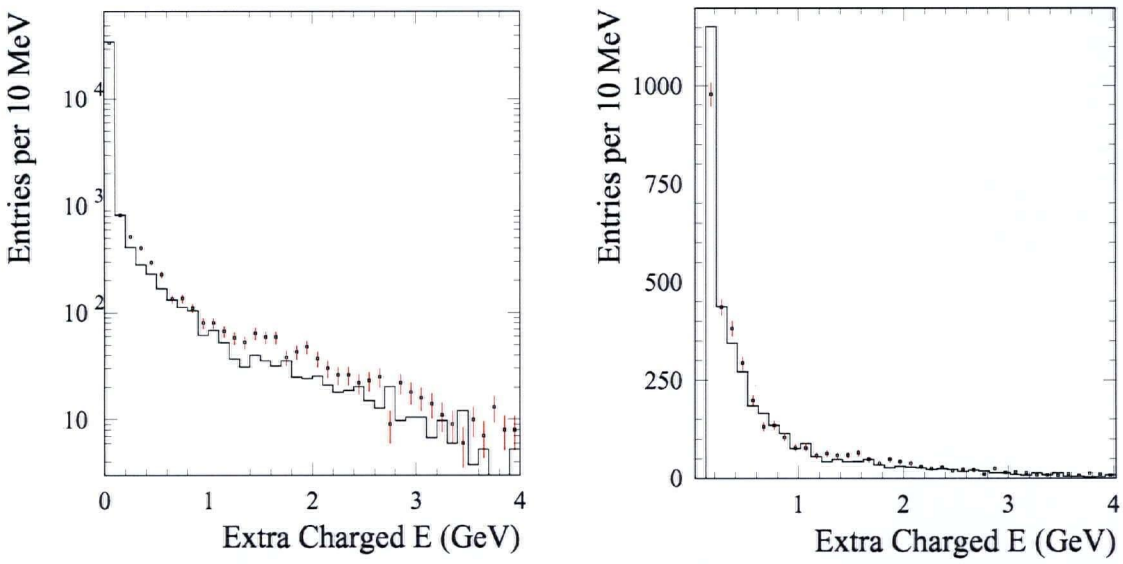

Figure 6-8: Extra charged energy in the event. The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale. (b) Linear scale for $E_{Extra-chg} > 20$MeV.

	MC	Data	Correction factor
Average Extra Charged Energy (MeV)	53.2	73.7	1.4

Table 6-4: Data/MC comparison for extra charged energy in the event.

6.3 Extra Total Energy

The extra total energy is the sum of extra neutral energy and extra charged energy in the event. **Figure 6-9** and Table 6-5 show the Data/MC comparison of extra total energy in this analysis. As you can see, the MC underestimates the amount of total extra energy by a factor of 37%. By comparing the Data/MC agreement of extra neutral energy and extra charged energy in previous sections, we can see the largest effect in Data/MC disagreement in extra total energy comes from the disagreement in extra charged energy. Requiring there to be no extra charged tracks in the event could eliminate this difference. This practice is discussed in the next section.

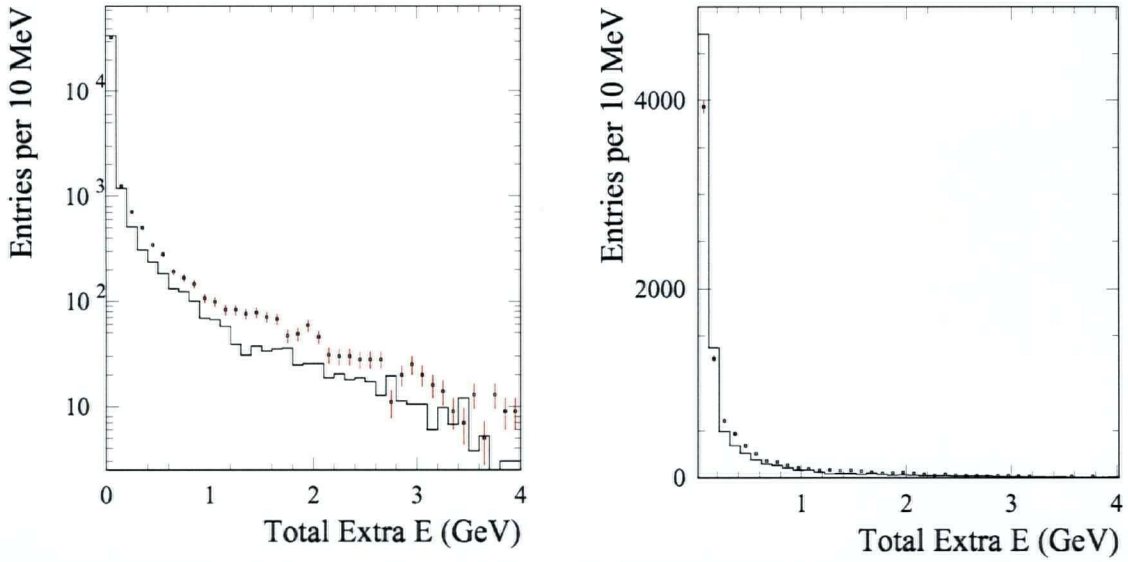

Figure 6-9: Extra total energy in the event. The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale. (b) Linear scale for $E_{Extra-total} > 20$MeV.

	MC	Data	Correction factor
Average Extra Total Energy (MeV)	60.4	94.9	1.6

Table 6-5: Data-MC comparison for extra total energy in the event.

6.4 Extra Energy for Events with No Extra Charged Tracks

As we demonstrated in section 6.2.1, about 90% of the events do not have any extra charged tracks in them. Now we investigate if we look only at these events, would we get better agreement between data and MC when looking at extra energy in the event?

Obviously, in this case, we do not have any extra charged energy, since we do not have any extra charged tracks, and hence, the total extra energy in the event is equal to the extra neutral energy in the event. The results are presented in Figure 6-10 and Table 6-6. As you can see, the MC underestimates the amount of extra energy by a factor of 17% which is much better than the factor presented in the previous section (37%), but not better than the results in the extra neutral energy (section 6.1). This implies that most of the disagreement between data and MC in section 6.1 comes from poor simulation of hadronic showers of signal tracks, rather than background tracks. Table

6-6 also provides the information of what percentage of these events have extra energy greater than 250 MeV. This value is used as a cut to extract the signal in the $B^{\pm} \to K^{\pm}\upsilon\overline{\upsilon}$ study [23]. We can see that, the MC underestimates the data by a factor of 89%.

Table 6-7 provides a comparison of the extra energy near the charged tracks and near the gamma. The MC underestimates the amount of hadronic showers near the charged tracks by a factor of 73% and underestimates the EM shower near the high energy photon by a factor of 29%.

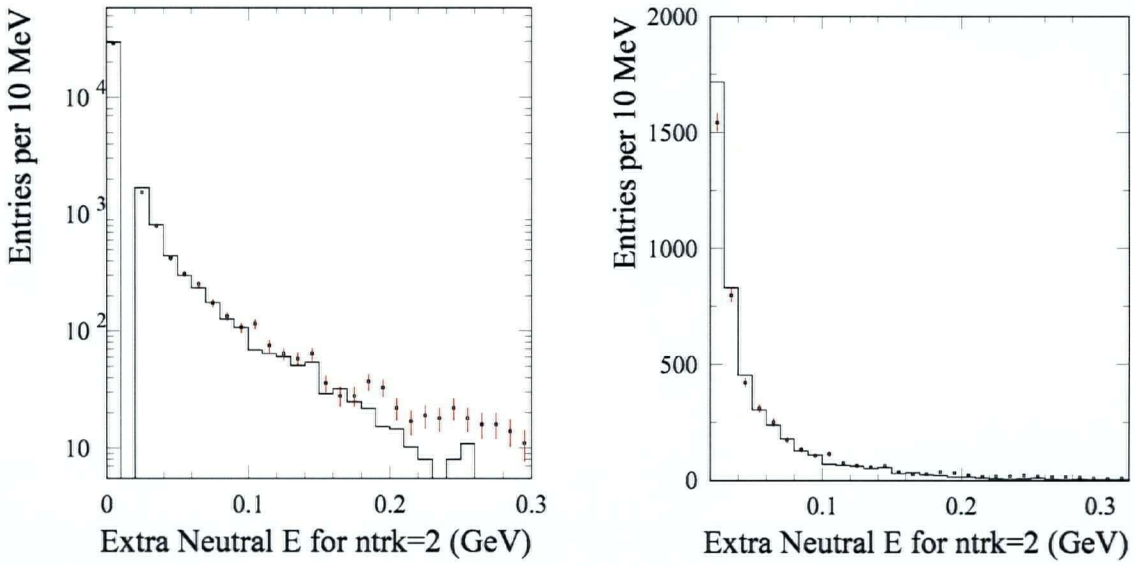

Figure 6-10: These plots are the same plots as in Figure 6-9, but only for the events with N_{trk}=2. The dotted points are Data; the histogram is MC that has been normalized to Data luminosity. (a) Logarithmic scale of total extra energy in the event. (b) Linear scale of total extra energy in the event for $E_{Extra-total} > 20$ MeV

	MC	Data	Correction factor
Average Extra Total Energy (MeV)	6.7	8.0	1.2
Percentage of events with extra energy >250MeV	0.2%	1.6%	9.0

Table 6-6: Data/MC comparison for extra energy in the event where N_{trk}=2.

		MC	Data	Correction factor
For $\psi_{trk} < 1$ Rad.	Energy deposited (MeV)/charged track	3.5	13.1	3.7
For $\psi_{gamma} < 1$ Rad.	Energy deposited(MeV)/gamma	1.7	2.4	1.4

Table 6-7: Data/MC comparison of the total extra energy near the charged tracks and the gamma in the events with no extra charged tracks.

6.5 Studying the Cuts Used in BaBar Document #633

BaBar Analysis Document (BAD) #633[31] discusses the choice of selection criteria for charged tracks, neutral clusters and composite particles. The goal is to optimize the reconstruction of visible energy while maintaining good agreement between data and simulation. This section provides an overview of the cuts mentioned in this document and we then study the effectiveness of these cuts by studying the extra energy after implementing them on our signal events.

6.5.1 Neutral Cluster Selection

A number of cluster variables were studied:

1. Number of crystals N_{Crys} in the cluster
2. Energy of the cluster E_{Clus} in the lab frame
3. Cluster lateral moment $latMom$
4. θ_{Clus} of the cluster in the lab frame
5. Angle between the position of the cluster and the impact point of the nearest charged track at the EMC surface; the 3-d angle is given by: $\Delta\alpha = Cos^{-1}[Cos\theta_{cl}Cos\theta_{tr} + Sin\theta_{cl}Sin\theta_{tr}Cos(\varphi_{cl} - \varphi_{tr})]$. In addition, $\Delta\theta$ and $\Delta\varphi$ are defined as the track minus cluster difference. Note that $\Delta\varphi$ is multiplied by the charge of the track closest to the cluster; this will make the distribution of $\Delta\varphi$ asymmetric about zero.
6. Zernike moments A20 and A42[21].

[21] The Zernike moment A_{nm} is calculated using the energy E_i and location (ρ_i, φ_i) of crystals with respect to the shower centroid. The location is defined in a cylindrical coordinate system with the z axis

The philosophy behind choosing the cuts was to remove regions where the cluster yield is dominated by background photons, and to try to avoid regions with large data–Monte Carlo differences. The following cuts were chosen:

1. $N_{Crys} > 2$
2. $E_{Clus} > 50$ MeV
3. $latMom < 0.6$ (LAT is a shower shape variable. it is the width of a cluster in the EMC, normalized to be between 0 and 1[22])
4. $0.32 < \theta_{Clus} < 2.44$
5. $\Delta\alpha > 0.08$ for tracks that do not pass the electron likelihood selector. No cut is made on this variable for electrons.

No cuts are made on the Zernike moments, as it was found that they added no new information. It should be noted that the common practice of requiring $latMom > \varepsilon$, where ε is a small value (e.g. 0.01) has the effect of cutting out clusters with less than 3 crystals, since $latMom = 0$ when $N_{Crys} < 3$. We choose here to place the cut more transparently on N_{Crys}.

6.5.2 Charged Track Selection

The optimum cuts for the track selection follow several guidelines:

1. To reconstruct the missing neutrino (in events containing a neutrino) with optimal resolution, one wants to select in an event as many good tracks as possible.

---

running from the beam spot to the centroid, where $\rho_i = r_i / R_0$ and $R_0 = 15$ cm. $A_{nm} = \sum_i^n (E_i / E) \times f_{nm}(\rho_i) e^{-im\varphi}$, where the sum includes only crystals with $\rho_i < 1$, and E is the total energy in the cluster. The Zernike functions are $f_{nm}(\rho) = \sum_{s=0}^{(n-m)/2} [(-1)^s (n-s)! \rho^{(n-s)} / s!((n-m)/2-s)!((n+m)/2-s)!]$ with $m \le n$ and $(n-m)$ even. Studies indicate that A_{42} provides good separation between hadronic and electromagnetic showers when used in conjunction with LAT.

[22] $LAT = \sum_{i=3}^{n} E_i r_i^2 / (E_1 r_0^2 + E_2 r_0^2 + \sum_{i=3}^{n} E_i r_i^2)$, where the n crystals in the EMC cluster are ranked in order of energy deposited in that crystal, E_i, and $r_0 = 5$ cm is the average distance between crystal centers. r_i is the distance between crystal i and the cluster centroid calculated from an energy-weighted average of the n crystals.

2. One would like to minimize biases by rejecting tracks due to misreconstruction in the tracking algorithm (e.g. loopers, ghosts).
3. In order to find the optimum cuts we distinguish SVT-only tracks, DCH-only tracks and SVT+DCH tracks.

Here, we summarize the final set of selection cuts applied to the charged track sample. The subsequent sections detail the criteria for loopers and ghosts.

We keep tracks passing the following cut criteria:

1. $0.41 < \theta_{Lab} < 2.54$
2. $P_{Lab} < 10$ GeV/c
3. $P_{t,Lab} > 0.06$ GeV/c
4. $|DOCA| < 1.5$ cm (DOCA is the distance between the decay point of the mother particle and the origin, in the xy plane.)
5. $|z_0| < 5$ cm (z_0 is the distance between the decay point of the mother particle and the origin, along the z axis.)
6. $P_{t,Lab} < 0.2$ GeV/c for SVT-only tracks (P_t is the projection of the momentum of the track on the xy plane)
7. No looper tracks (see below)
8. No ghosts (see below)

6.5.2.1 Looper rejection

A looper is produced by a particle with small momentum looping inside the tracking system. The tracking reconstruction algorithm is not able to identify this as belonging to one single track. In this case, it tends to reconstruct several track candidates, the number of which depending on the ratio of transverse to longitudinal momentum. These additional track candidates will have almost the same transverse momentum. However, since the detector has a restricted size additional track candidates from loopers will mainly occur for $\theta_{Lab} \approx \pi/2$. The large majority of looper candidates are SVT-only tracks.

We identify looper candidates by requiring for any pair of charged tracks: $P_{t,Lab} < 0.25$ GeV/c, $|Cos\theta_{Lab,Trk}| < 0.2$, and $|\Delta P_{t,Lab}| = |P^i_{t,Lab} - P^j_{t,Lab}| < 0.12$ GeV/c. We then calculate $\Delta\theta_{Lab} = \theta^i_{Lab} - \theta^j_{Lab}$ and $\Delta\varphi_{Lab} = \varphi^i_{Lab} - \varphi^j_{Lab}$. If the particle loops exactly one time then the tracking algorithm will find two tracks going back-to-back with opposite charges assigned. If the particle loops more than one time then there is at least one additional track with the same charge assigned which will have similar θ_{Lab} and φ_{Lab}. Hence, there are two possibilities for $\Delta\theta_{Lab}$ and $\Delta\varphi_{Lab}$. For the opposite charge assigned we redefine $\Delta\theta_{Lab}$ and $\Delta\varphi_{Lab}$ in the following way: $\Delta\theta_{Lab} = \pi - \Delta\theta_{Lab}$, $\Delta\varphi_{Lab} = \pi - \Delta\varphi_{Lab}$ if $\Delta\varphi_{Lab} > \pi/2$ and $\Delta\varphi_{Lab} = \pi + \Delta\varphi_{Lab}$ if $\Delta\varphi_{Lab} < -\pi/2$. We

then apply the cut $|\Delta\varphi_{Lab}| < 0.18$ for same charge tracks and $|\Delta\varphi_{Lab}| < 0.16$ for opposite charged ones.

Looper candidates are then defined as those track candidates where at least one pair of tracks have been found passing the additional cuts: $|\Delta\theta_{Lab}| < 0.18$ (for opposite charges) and $|\Delta\theta_{Lab}| < 0.20$ (for equal charges), respectively.

For a set of track candidates marked to belong to the same looper we retain the track with the smallest $|z_0|$ value.

6.5.2.2 Ghost rejection

Ghost tracks are the ones that do not correspond to a particle in the detector. We distinguish two kinds of so-called ghost tracks. The first class is SVT-only tracks with large momenta ($P_{t,Lab} > 0.2$ GeV/c), which we consider them as an artificial product of the tracking search algorithm (otherwise for large momenta tracks, we should have hits in the DCH as well).

Hence, SVT-only tracks with $P_{t,Lab} > 0.2$ GeV/c are removed from the analysis.

The second category belongs to cases where the tracking algorithm has reconstructed two tracks out of one. One may think of various possibilities of how this may happen: additional random hits, decays in flight, etc. This case can be identified by searching for track pairs with close $P_{t,Lab}$, θ_{Lab} and φ_{Lab}. For such track candidates, we expect that the number of hits per track is smaller than the average number of hits per track.

It turns out that the majority of DCH-ghost candidates occur mainly at low momenta: 87% of the candidates are found for $P_{t,Lab} < 0.35$ GeV/c. We select candidates for DCH-ghosts by requiring: $|\Delta P_{t,Lab}| < 0.15$GeV/c, $|\Delta\theta_{Lab}| < 0.3$ and $|\Delta\varphi_{Lab}| < 0.3$. One of the two tracks has to be retained. We reject the track with the smaller number of DCH-hits. However, this is not a sufficient criterion. There is a significant enhancement around 20 DCH-hits. This fits to our expectation that the typical number of 40 DCH-hits is shared between two track candidates. There is a second enhancement around 40 hits. Consequently, we consider as DCH-ghost candidates only those tracks with a number of DCH-hits smaller than the cut $(nDCH(track1) < 45 - nDCH(track2))$. As a cross check, we compare the fit probability corresponding to the χ^2 of the track fit and find that this selection criterion does indeed choose, on a statistical basis, the track with the better fit probability. In addition, we compare the measured kinematic variables with the MC truth values and find also that the bias is smaller for the track with the larger number of DCH-hits.

Extra Energy Study after Cuts on Neutral Clusters and Charged Tracks

6.6 Extra Neutral Energy

Figure 6-11 and Table 6-8 present Data/MC comparisons for extra neutral energy after implementing the cuts mentioned in the previous section. As can be seen the data underestimates the $E_{Extra-neutral}$ by a factor of 33%. We can see that in this part implementing the cuts has worsened the agreement between data and MC (Table 6-1). But it should be pointed out that the difference between data and MC is the same as what we had before implementing the cuts (1.3 MeV) and also note that after implementing the cuts, the fraction of events with $E_{Extra-neutral} \leq 20$ MeV has increased which is an improvement.

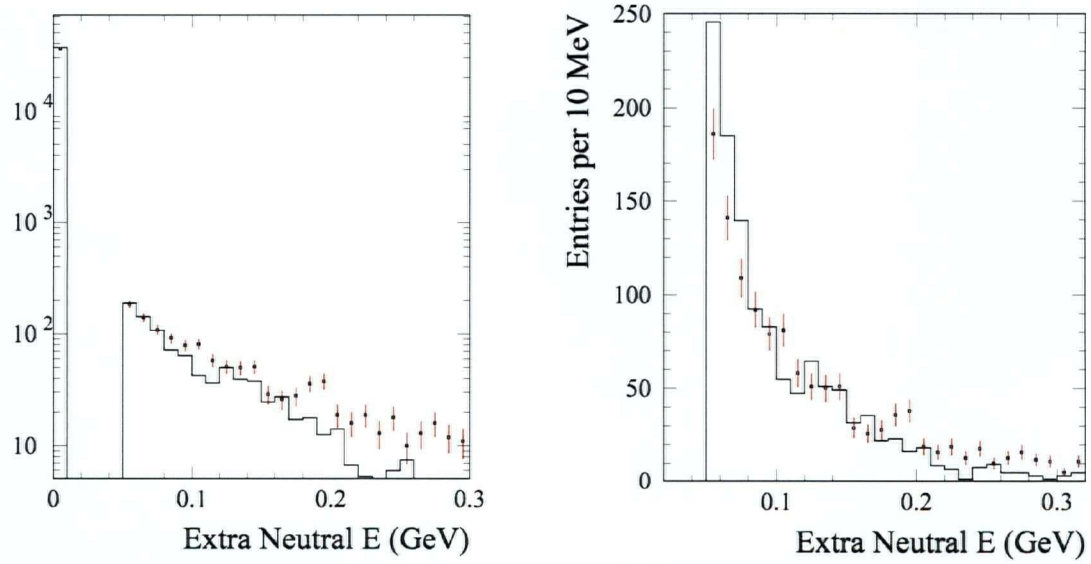

Figure 6-11: Extra neutral energy in the event after applying the cuts from BAD #633. The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale. (b) Linear scale for $E_{Extra-neutral} > 20$MeV.

	MC	Data	Correction factor
Average Extra Neutral Energy (MeV)	2.5	3.8	1.5
Fraction of Events with Extra Neutral Energy ≤20MeV	97.5%	95.5%	1.0

Table 6-8: A comparison between average extra neutral energy for events in data and MC after applying the cuts from BAD #633.

6.6.1 Neutral clusters near a charged track

Figure 6-12 shows that the MC still underestimates the number of clusters after applying the cuts. We can also see a considerable decrease in the number of clusters in data near the signal photon, but the peak near the signal photon, although smaller, still exists in the MC (Figure 6-12). Figure 6-13, Figure 6-14 and Figure 6-15 present the same set of plots as Figure 6-4, Figure 6-5 and Figure 6-6. The overall shape of variables other than those mentioned above does not seem to have been affected much by implementing the cuts. However, a big decrease in the total number of clusters is apparent after the cuts are implemented. The summary of some numerical comparisons between data and MC is provided in Table 6-9. We can see that the agreement is clearly worse than what we had before implementing the cuts (Table 6-2). It should be pointed out that applying the cuts has excluded the low energy clusters near gamma as one can conclude by comparing the last line of Table 6-9 with that of Table 6-2.

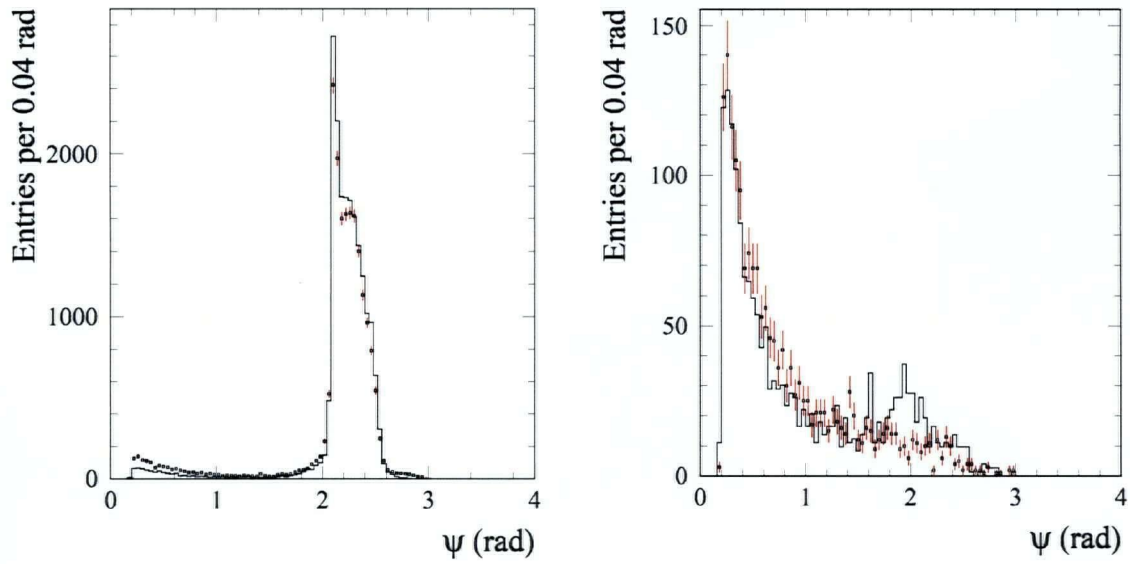

Figure 6-12: The dotted points are Data, and the histogram is MC that has been normalized to Data luminosity. (a) ψ distribution for all clusters after applying the cuts from BAD #633. (b) ψ distribution for all clusters except high-energy signal photon after applying the cuts from BAD #633.

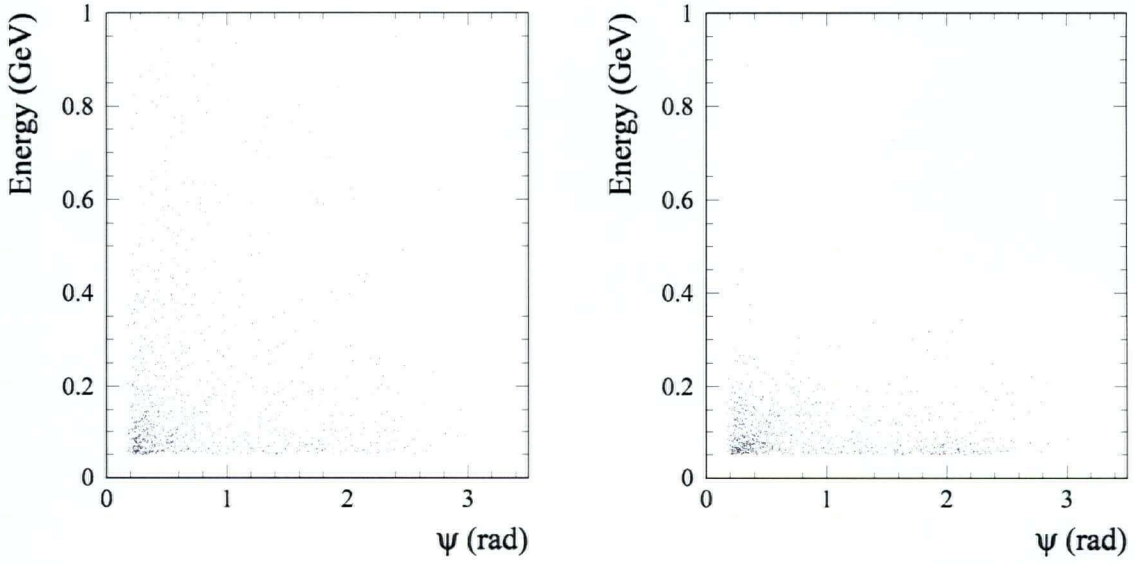

Figure 6-13: (a) The energy released at each cluster versus ψ_{trk} of the cluster for the Data after applying the cuts from BAD #633. (b) Energy released at each cluster versus ψ_{trk} of the cluster for the MC after applying the cuts from BAD #633.

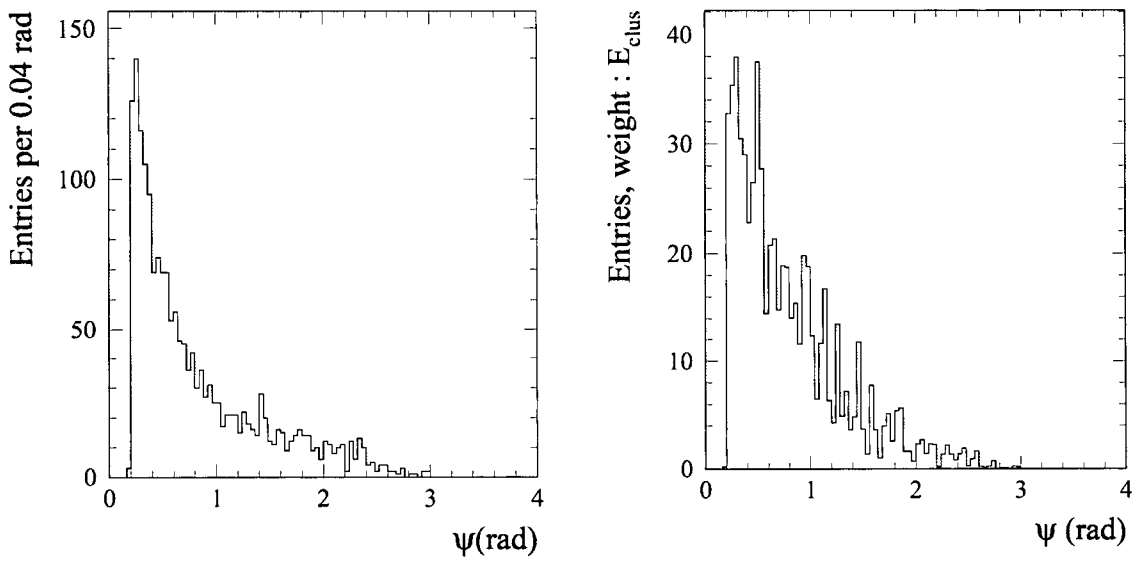

Figure 6-14: (a) ψ_{trk} distribution in data after applying the cuts from BAD #633. (b) ψ_{trk} distribution for data when each entry is weighted by the energy of the associated cluster after applying the cuts from BAD #633.

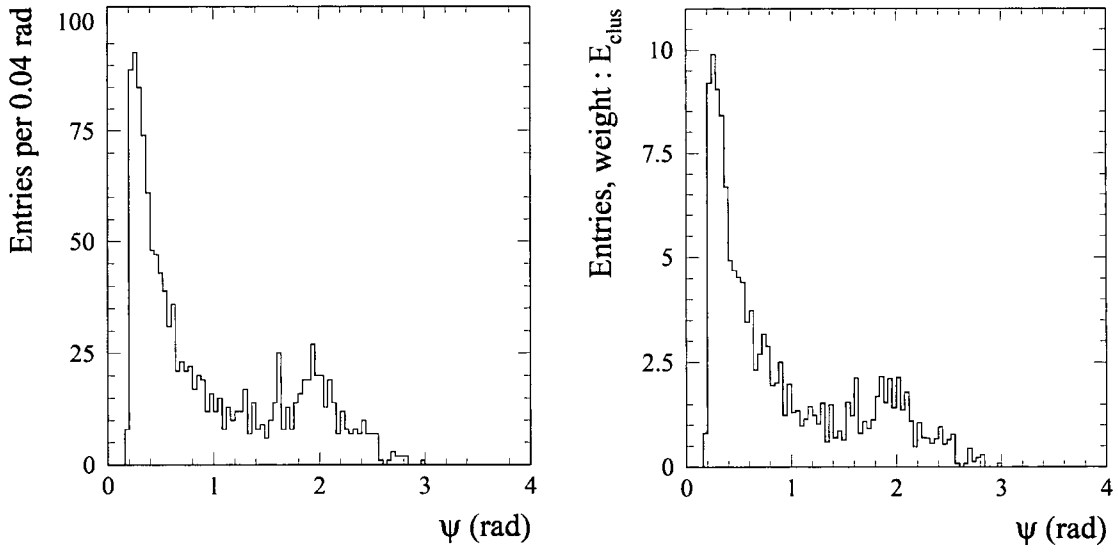

Figure 6-15: (a) ψ_{trk} distribution in MC after applying the cuts from BAD #633. (b) ψ_{trk} distribution in MC when each entry is weighted by the energy of the associated cluster after applying the cuts from BAD #633.

		MC	Data	Correction Factor
For $\psi < 1$ Rad.	# of clusters/charged track	0.016	0.034	2.1
	Energy deposited (MeV)/charged track	1.8	12.4	7.0
For $\psi_{gamma} < 1$ Rad.	Energy deposited(MeV)/gamma	0.5	1.2	2.4
	Energy deposited(MeV)/cluster	86.7	229.6	2.6

Table 6-9: Data/MC comparison for the number of clusters near a charged track and the high energy photon and the amount of energy deposited near them after applying the cuts from BAD #633.

6.7 Extra Charged Tracks and Energy

6.7.1 Extra Charged Tracks

Figure 6-16 and Table 6-10 demonstrate the number of extra charged tracks after the cuts are implemented. Here we can see that the MC still underestimates the number of extra charged tracks compared to data but the fraction of events with extra charged tracks have been reduced compared to the result before applying the cuts (Table 6-3), which is an improvement.

6.7.2 Extra Charged Energy

The results for data and MC are shown in Figure 6-17 and Table 6-11. We can see that there was much better agreement between data and MC before implementing the cuts (**Figure 6-8** and Table 6-4). Here the MC underestimates the amount of extra charged energy by a factor of 2.4.

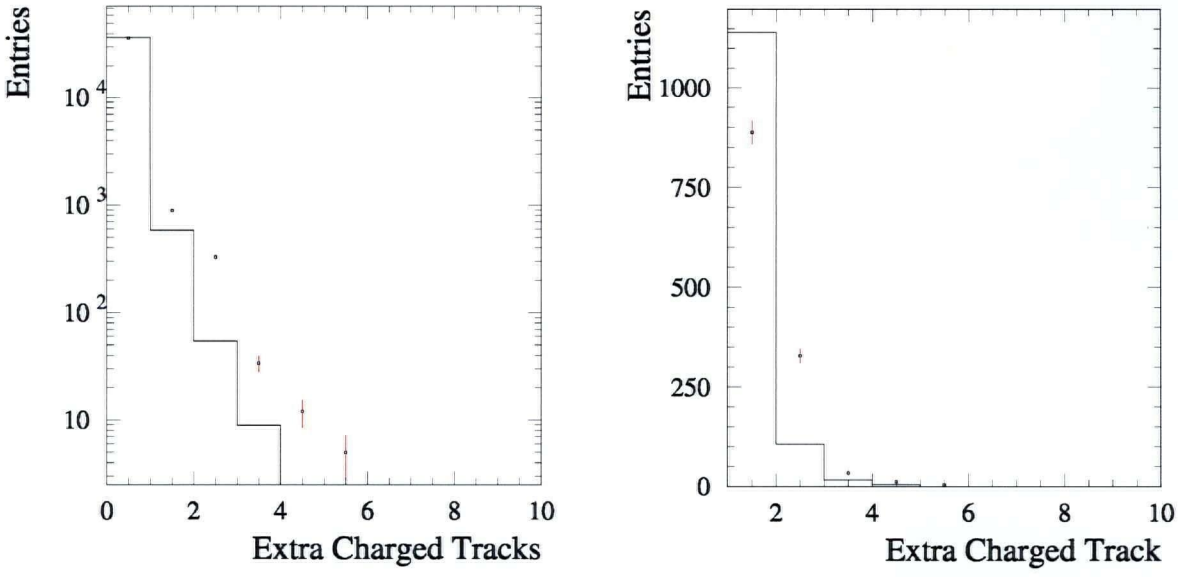

Figure 6-16: The points refer to data; the histogram is MC that has been normalized to Data luminosity. (a) Logarithmic scale of number of extra charged tracks in the event after applying the cuts from BAD #633. (b) Number of extra charged tracks in the event for $N_{Trk-extra} > 1$ after applying the cuts from BAD #633.

	MC	Data	Correction Factor
Percent of events with extra charged tracks	1.7%	3.3%	1.9

Table 6-10: A comparison between number of extra charged track for events in data and MC after applying the cuts from BAD #633.

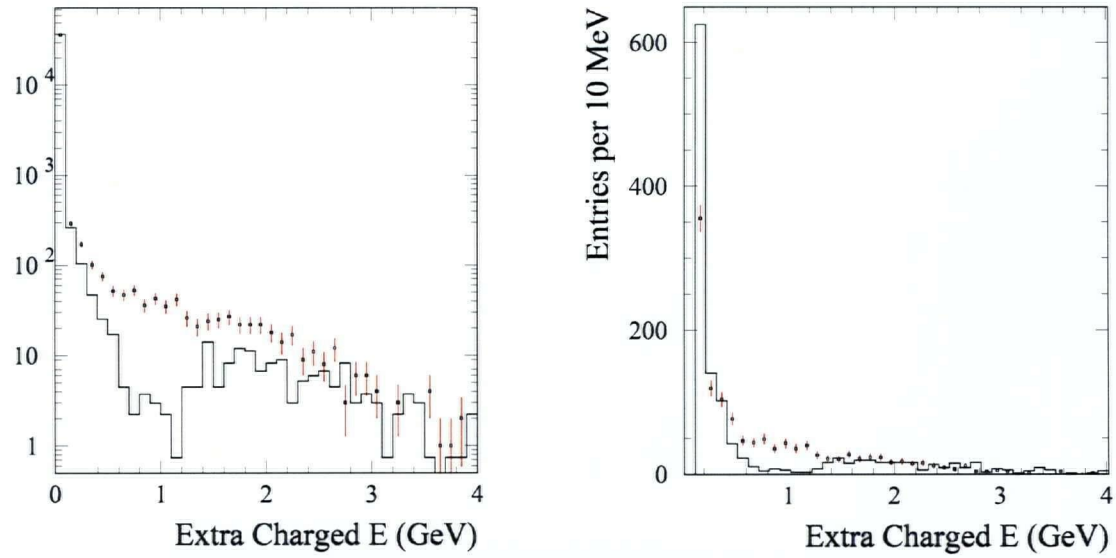

Figure 6-17: Extra charged energy in the event after applying the cuts from BAD #633. The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale. (b) Linear scale for $E_{Extra-charged} > 20$MeV.

	MC	Data	Correction factor
Average Extra Charged Energy (MeV)	11.2	26.8	2.4

Table 6-11: Data-MC comparison for extra charged energy in the event after applying the cuts from BAD #633.

6.8 Extra Total Energy

Figure 6-18 and Table 6-12 show the Data-MC comparison for this quantity after implementing the cuts mentioned in BAD #633. Once again we see that we had much better agreement before the cuts were applied (Table 6-9). Here the MC underestimated the amount of total extra energy by a factor of 3.1.

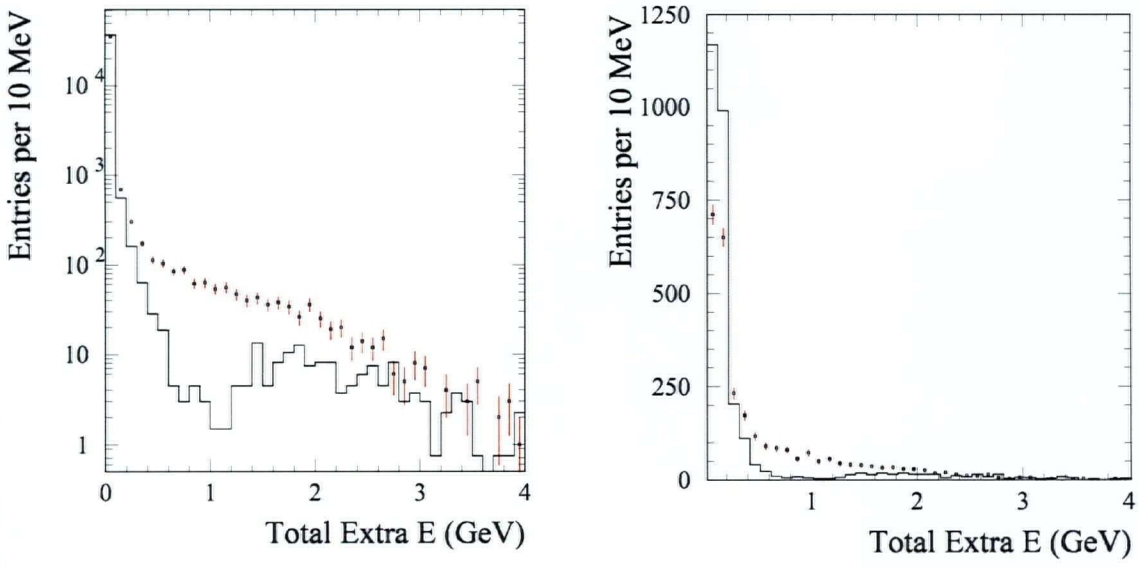

Figure 6-18: Extra total energy in the event after applying the cuts from BAD #633. The points refer to data; the histogram is MC that has been normalized to Data luminosity (i.e. number of data entries in each window). The goal is to compare the distribution shape in data versus MC. (a) Logarithmic scale. (b) Linear scale for $E_{Extra-total} > 20$MeV.

	MC	Data	Correction factor
Average Extra Total Energy (MeV)	13.9	42.9	3.1

Table 6-12: Data-MC comparison for extra total energy in the event after applying the cuts from BAD #633.

6.9 Extra Energy for Events with No Extra Charged Tracks

In this part we can also see that the agreement is not as good as the agreement before applying the cuts. Nevertheless, we have fewer events with extra energy greater than 250 MeV, which makes that cut more reliable in this respect. (Figure 6-19, Table 6-13 and Table 6-14).

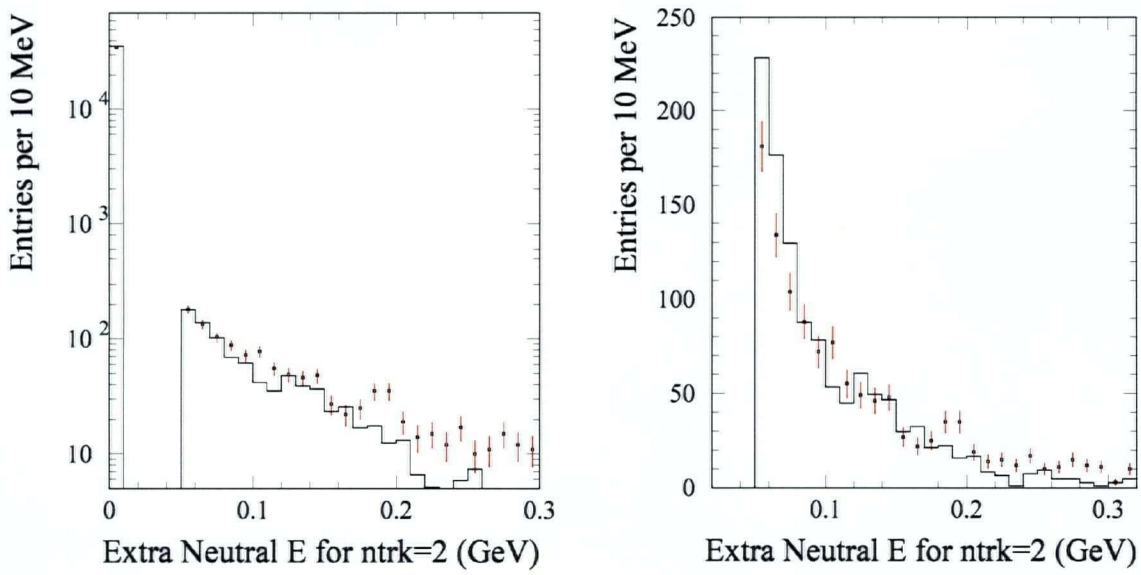

Figure 6-19: These plots are the same plots as in Figure 6-18, but only for the events with N_{trk}=2. The dotted points are Data, and the histogram is MC that has been normalized to Data luminosity. (a) Logarithmic scale of total extra energy in the event. (b) Linear scale of total extra energy in the event for $E_{Extra-total} > 20$ MeV.

	MC	Data	Correction factor
Average Extra Total Energy (MeV)	2.5	3.7	1.5
Percentage of events with extra energy >250MeV	0.1%	1.4%	12.9

Table 6-13: Data-MC comparison for extra energy in the event where N_{trk}=2 after applying the cuts from BAD #633.

		MC	Data	Correction factor
For $\psi_{trk} < 1$ Rad.	Energy deposited (MeV)/charged track	1.4	10.7	7.6
For $\psi_{gamma} < 1$ Rad.	Energy deposited(MeV)/gamma	0.5	1.1	2.2

Table 6-14: Data/MC comparison of the total extra energy near the charged tracks and the gamma in the events with no extra charged tracks after applying the cuts from BAD #633.

6.10 Conclusion

Table 6-15 provides the summary of all the tables in this chapter. As you can see, the best agreement between data and MC is when there are no extra charged tracks in the event. It should also be noted that for all of the cases, implementing the cuts mentioned in BAD #633[31] have worsened the agreement between data and MC.

		Before BAD #633 cuts			After BAD #633 cuts		
		MC	Data	Corr. Factor	MC	Data	Corr. Factor
Avg. $E_{Extra-neutral}$ (MeV)		7.0	8.3	1.2	2.5	3.8	1.5
Fraction of Events with $E_{Extra-nutral} \leq 20$MeV		86.9%	85.2%	1.0	97.5%	95.5%	1.0
ψ_{trk} <1 rad	#clus/chg trk	0.065	0.86	1.3	0.016	0.034	2.1
	$E_{Extra-nutral}$ /chg trk	4.0	15.0	3.8	1.8	12.4	7.0
ψ_{gamma} <1 rad.	$E_{Extra-nutral}$ (MeV)/gamma	1.9	2.7	1.4	0.5	1.2	2.4
	$E_{Extra-nutral.}$ (MeV)/cluster	41.1	62.4	1.5	86.7	229.6	2.6
% event with extra charged tracks		8.3%	10.1%	1.2	1.7%	3.3%	1.9
Avg. $E_{Extra-charged}$ (MeV)		53.2	73.7	1.4	11.2	26.8	2.4
Avg. $E_{Extra-total}$ (MeV)		60.4	94.9	1.6	13.9	42.9	3.1
Avg. $E_{Extra-total}$ (N_{trk}=2)		6.7	8.0	1.2	2.5	3.7	1.5
N_{trk}=2	ψ_{trk} <1 Rad. $E_{Extra-nutral}$ (MeV)/gamma	3.5	13.1	3.7	1.4	10.7	7.6
	ψ_{gamma} <1 Rad. $E_{Extra-nutral}$ (MeV)/cluster	1.7	2.4	1.4	0.5	1.1	2.2
% of events with N_{trk}=2 and $E_{Extra-total}$ >250MeV		0.2%	1.6%	9.0	0.1%	1.4%	12.9

Table 6-15: Summary of the results obtained in this chapter

7

Studying the Kaon Identification System Efficiency

7.1 Introduction

In this chapter, we will extract the efficiency of the kaon identification system using the $\varphi\gamma$ control sample for data and MC. The underlying difference between this study and the left over energy study is that here the fact that the φ decays into charged kaons plays the most important role, whereas in the previous chapter, we could have used any mode of decay for the φ, as long as we could get a good estimate of the energy of the daughter particles. As previously mentioned (Chapter 5) for this study, we only require one of the kaons to pass a tight KID selector, and we assume the other track is a kaon if it makes a nice φ mass peak when combined with the first one. The kaon identification efficiency (ε) is defined as the fraction of true kaons identified by a kaon selection (Section 7.4). In this study we extract the true number of kaons in the sample from the number of φ mesons obtained from a fit to the mass peak.

We extract the efficiency of the system for both K^+ and K^- since in general we might have different efficiencies for identifying positively charged kaons and negatively charged ones (Section 7.3). For each charge we also check how much the efficiency changes if we require the tracks to make it to the DIRC i.e. if we only work with the tracks that actually have some DIRC information associated with them to help with the PID (Particle Identification)[23]. Table 7-1 summarizes the notation used in this study for these different efficiencies.

[23] Although the kaon selection system uses the DIRC information, it does not require the tracks to make it to the DIRC. For the ones that don't, the selection system selects the tracks just based on SVT and DCH information.

	Makes it to the DIRC	Doesn't make it to the DIRC
K^+	$\varepsilon_{K^+,D}$	$\varepsilon_{K^+,NoD}$
K^-	$\varepsilon_{K^-,D}$	$\varepsilon_{K^-,NoD}$

Table 7-1: Different efficiencies studied in this study

7.2 Detection of Kaons in BaBar[26]

The Kaon selection in BaBar is mainly based on the information that comes from the silicon-vertex tracker (SVT), the drift chamber (DCH) and the internally reflecting Cherenkov detector (DIRC). From the truncated-mean dE/dX measurements a better than 2σ separation of pions and kaons is possible up to about 0.6 GeV/c particle momentum in the SVT and 0.7 GeV/c in the DCH (see the dE/dX distribution of pions and kaons for the SVT in Figure 7-1 (left) and for the DCH in Figure 7-1 (right)). For momenta above 1.5 GeV/c the DCH provides about 2σ dE/dX separation between kaons and pions (due to the relativistic rise in the Bethe-Bloch formula[24]). Figure 7-2 shows the distribution of SVT and DCH dE/dX versus momentum for a generic B-meson decay Monte Carlo (MC).

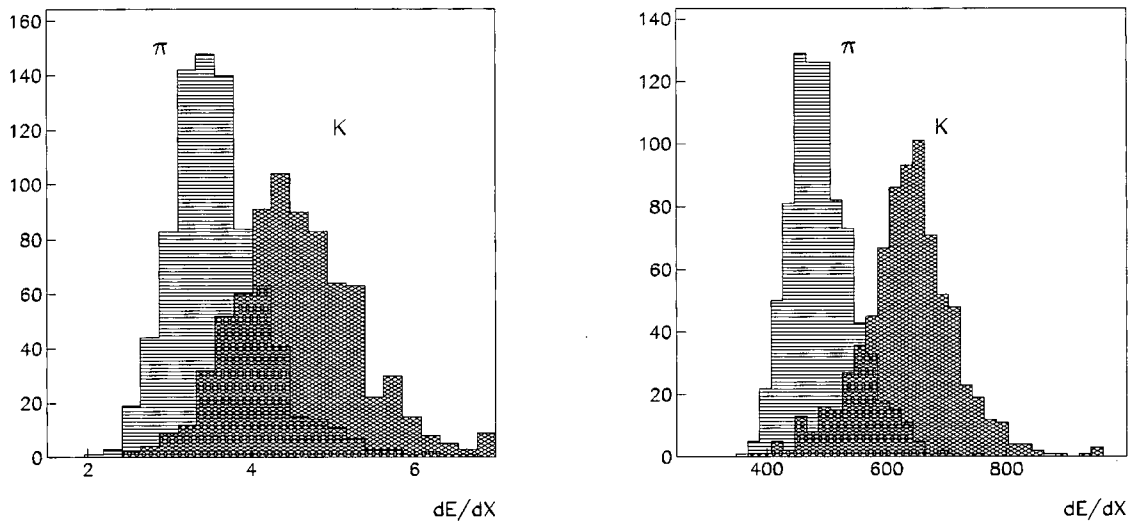

Figure 7-1: Left: SVT truncated-mean dE/dX distribution for pions and kaons in the momentum range from 0.5 to 0.6 GeV/c kinematically identified in a very pure D^0 decay sample. Right: DCH dE/dX distribution from the same data sample as above in the momentum range from 0.6 to 0.7 GeV/c.

[24] Describes the energy loss of a charged particles in matter[27].

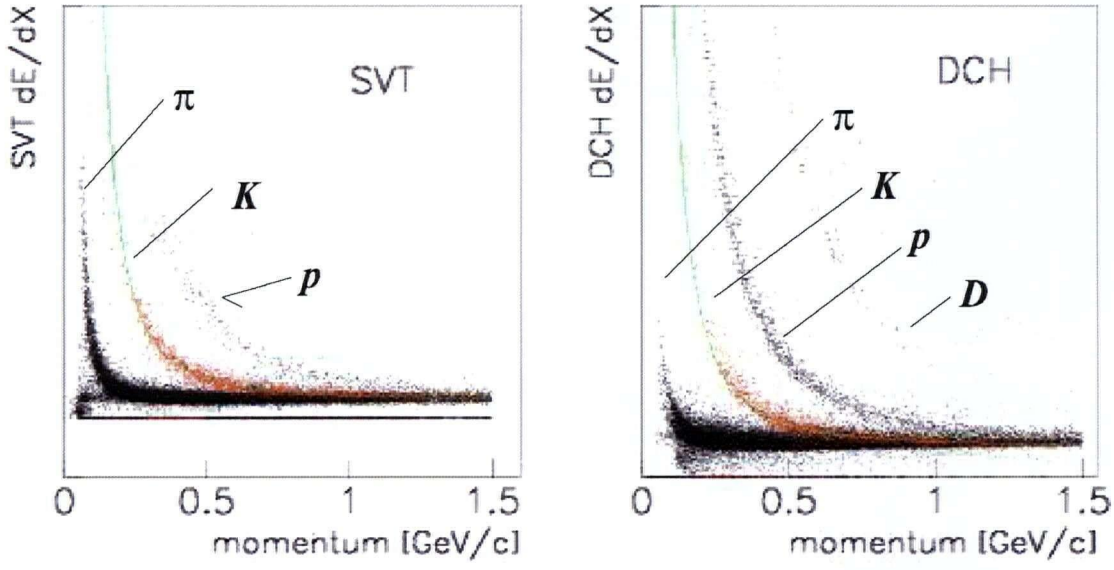

Figure 7-2: Monte Carlo dE/dX versus particle momentum distribution for the SVT (left) and the DCH (right).

Kaons emit Cherenkov light in the fused silica bars of the DIRC (index of refraction n =1.473) for momenta above the threshold:

$$p > \frac{m}{\sqrt{n^2 - 1}}$$

Equation 7-1

Where p and m are the particle momentum and mass, respectively. The central value of the Cherenkov emission angle (wavelength $\lambda \approx$ 400 nm) with respect to the track, θ_C, follows:

$$\cos\theta_C = \frac{1}{n\beta} \text{ With } \beta = p/E$$

Equation 7-2

The number of photons produced for a fixed particle path in the fused silica follows Poisson statistics with a central value dependent on the particle type, charge, momentum, polar angle and bar number (position). This expected number of photons is stored in a lookup table as a function of these parameters. The behavior of the Cherenkov angle as a function of momentum is demonstrated with a generic B-decay Monte Carlo sample in Figure 7-3.

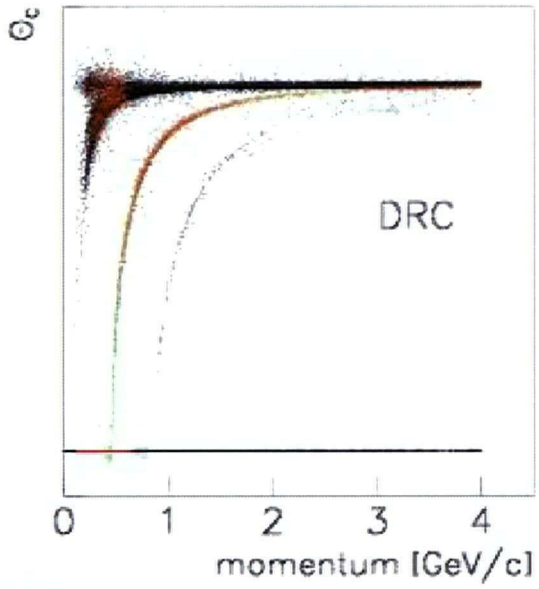

Figure 7-3: Distribution of the Cherenkov angle versus momentum for the different particle types reconstructed from a generic B-meson Monte Carlo sample.

Figure 7-4 shows the Gaussian width of the θ_C residual distribution for pions from a D^* control sample for different momentum intervals. It is parameterized by a polynomial and translated to a separation plot in Figure 7-4 (right). Since the resolution is measured with kinematically identified pions only, the observed 2.5σ separation at 4GeV/c is somewhat optimistic, while the realistic value based on pions and kaons is about 2σ.

The momentum range of the kaon sample in this study is 0.6-5.5 GeV/c.

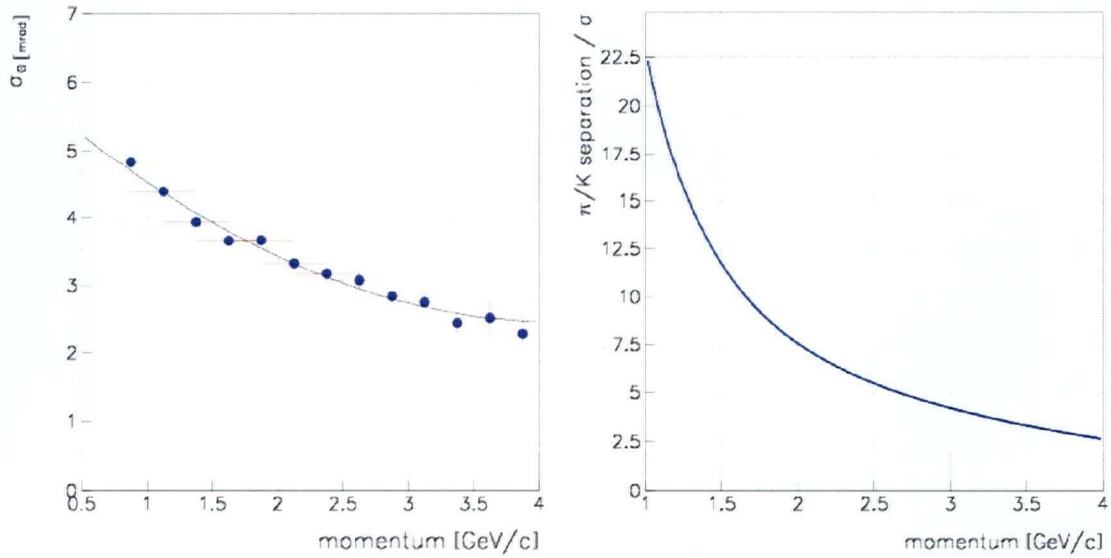

Figure 7-4: Cherenkov angle resolution for pions in a control sample $D^{*+} \to \pi^+ D^0, D^0 \to \pi K$. The data points are fit to a second order polynomial for an estimate of the separation in units of standard deviations. They are shown on the right.

7.3 Some details about charged Kaons

The signature of charged kaons in the different parts of the detector is complicated because they can decay or interact with the material of the detector. Furthermore, the hadronic interactions have different rates for the two charged modes.

Almost no charged kaons decay before the EMC, but it is important to notice that about 94% of the ones that do, result in a single charged track, which by following the initial kaon direction may not be distinguished from the original kaon. A decay-in-flight between the SVT and the DCH can cause ambiguous answers concerning particle identification of kaons versus muons or pions. Here, an algorithm, which is sensitive to kinks in the particle track, is provided in support of Kaon identification at low momenta[33]. It is not implemented in the selector directly, since the likelihood based selection already shows sensitivity to that effect when compared to the kink finder results.

Kaons that undergo $K^{\pm}N$ scattering before the DIRC can be misidentified. In fact not all will be lost after interacting with the detector, since elastically scattered kaons are still detectable to some extent. Also, some secondaries are misidentified as Kaons.

The difference in the hadronic interaction probability can lead to an asymmetry in the identification of the two charged types. We give an estimate of the asymmetry in the identification of the two charge types. The momentum spectra of positively charged kaons from a generic B meson decay Monte Carlo generator are displayed in Figure 7-5. This spectrum is folded with the ratio of the cross sections (interaction probability) of the two charges to arrive at the altered spectrum for negative kaons. It is superimposed on the previous plot in Figure 7-5.

The plot suggests that the effect due to detector material, even though it is observed at momenta around 1 GeV/c, is negligible. In the validation of the Kaon selectors' algorithm induced charge asymmetries are carefully monitored.

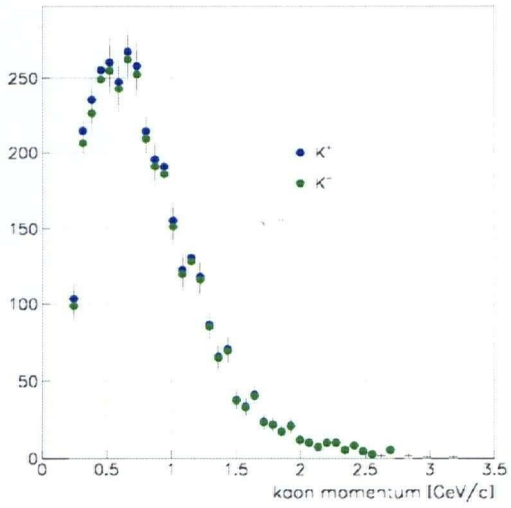

Figure 7-5: Lab momentum spectrum of positive and negative kaons from B mesons generated generically (not reconstructed). The negative Kaon spectrum is derived by multiplying the positive Kaon spectrum with the ratio of the KN total cross sections.

7.4 Principle of the Likelihood Selection

As an example let us consider the Cherenkov angle θ_C distribution for pions and kaons with respect to the expected angle, θ_π and θ_K. We assume that the distribution follows a Gaussian, and that the error, σ_{θ_C}, is well known. We define the pull:

$$\Delta_i = \frac{\theta_C - \theta_i}{\sigma_{\theta_C}}, \quad i = K, \pi$$

Equation 7-3

Δ_K is plotted in Figure 7-6 (left) for some momentum interval; for a kaon sample, Δ_K is centered at zero. The central value for Δ_π is shifted to a higher value, which corresponds to the separation between the mean values $\theta_\pi - \theta_K$. The probability density function is a Gaussian function g_i:

$$g_i = \frac{1}{\sqrt{2\pi}\sigma_{\theta_C}} e^{-0.5.\Delta_i^2}$$

Equation 7-4

To separate pions from kaons one can choose a ratio criterion

$$g_K / g_\pi > r, \quad \text{(i.e. } g_K > g_\pi \text{ for } r = 1\text{)}$$

Equation 7-5

Due to the overlap between the measured distributions for pions and kaons a certain fraction of pions will survive this selection. We define:

Kaon-identification efficiency ε is the fraction of kaons identified out of a pure kaon sample by a kaon selection.

Pion-misidentification rate κ is the fraction of pions from a pure pion sample that is selected by a kaon criterion.

The resulting kaon efficiency and pion misidentification for the criterion r = 1 is shown in Figure 7-6(right) - a fast degradation towards higher momenta is due to the fast reduction of the separation in Figure 7-4 (right). The picture in Figure 7-6 is idealized. Tails in the distributions lead to deviations. On the other hand a selection based on likelihood ratios only tests the closeness to the expected curve and hence smoothes out the tails or the asymmetric distributions. The analyst is able to maintain a certain efficiency or misidentification level by defining the (momentum and polar angle dependent) values of r [25].

[25] The log likelihood difference $\ln(g_K) - \ln(g_\pi)$ is a more appropriate quantity to tune the values for r.

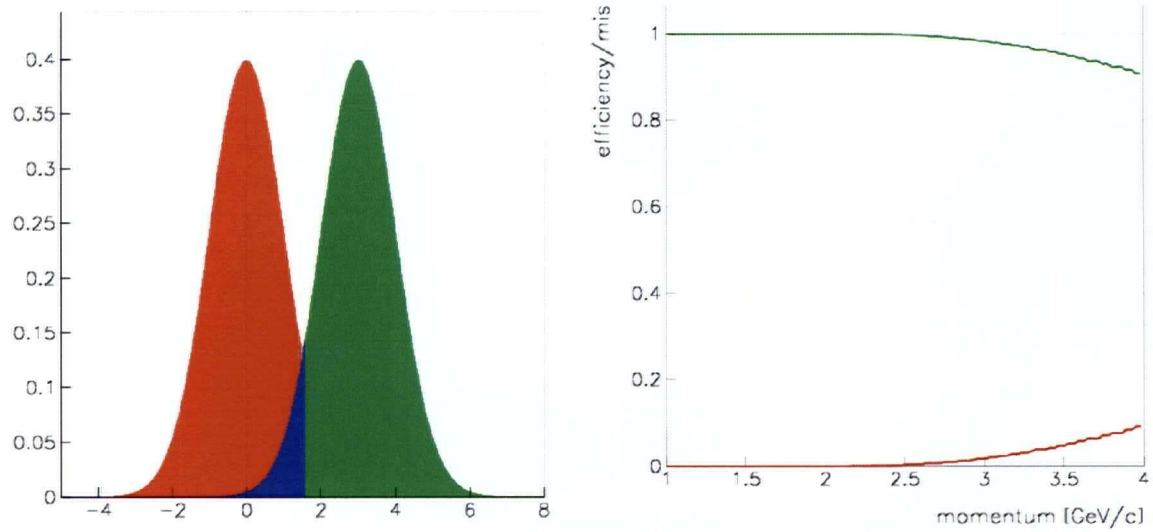

Figure 7-6: Assuming a Gaussian distribution of the Cherenkov angle around its mean (left), one can calculate the number of selected kaons and pions by a certain cut between the distributions of the two particle types. For an equal number of pions and kaons the kaon-identification efficiency and the pion-misidentification rate for the choice r = 1 is shown in the right plot.

7.5 Selector Families

There are 3 different families of selectors designed for Kaon identification:

- *PIDKaonSMSSelector*
- *PIDKaonNNSelector*
- *PIDKaonLHSelector*

There are 5 different modes based on tightness of the selection criteria for each Selector:

- *Very tight*
- *Tight*
- *Loose*
- *Very Loose*
- *Not a Pion*

All three selectors use likelihood calculations to distinguish between different particle types, mainly Kaons, Pions and Protons (the three "stable" charged hadron types).

For the two detectors, the SVT and the DCH, the PDF (the Probability Density Function) is a Gaussian distribution, where the mean is given by the dE/dx calibration, but the SMS uses a different parameterization of the errors on the DCH likelihood.

The momentum dependent relative error is provided by the calibration for both subsystems. The LH selectors use the beta-generated likelihoods for the SVT and the DCH. The selectors use different likelihoods for the DIRC: the *PIDKaonSMSSelector* uses a combination of a Poissonian and Gaussian derived from the fit results of the DIRC maximum likelihood fit [26]. The *PIDKaonNNSelector* uses the global likelihood[34],[35] which is a product of the individual photon likelihoods without an intermediate fit to a common Cherenkov angle and time distribution. Both approaches give comparable performances concerning the efficiency and misidentification for hadrons. The LH selector on the other hand, takes the Cherenkov angle and the number of photons and calculates a new likelihood, using an MC generated lookup table, which returns the correct probability for each particle type. This is the same information as is used in the SMS selector, but SMS makes a series of cuts to avoid non-Gaussian or non-Poisson regions.

The following subsection provides a brief discussion on the difference between SMS and LH selector. We will discuss the neural net selector afterward.

7.5.1 The PIDKaonSMSSelector and PIDKaonLHSelector

These two selectors are based on likelihood ratios between different particle types, but there is a basic difference in the "philosophy" of how these two selectors identify the "kaonness" of a track. Figure 7-7 shows the schematic view of pion and kaon likelihoods for a fixed momentum. The SMS selector is designed so that it will keep the Pion misidentification rate (κ) at a constant level for different momentum ranges. That is why we will see later on that the Kaon identification efficiency (ε') drops as the momentum increases. This happens because the separation between the two likelihoods in Figure 7-7 decreases as the momentum increases, i.e. the two Gaussians overlap more and more the higher the momentum. The two Gaussians are considered to be normalized. ε' denotes the efficiency of kaon identification for when the pion misidentification (κ) is kept constant as the momentum changes (as in the SMS selector). In this case, all tracks to the right side of the SMS line fail the kaon ID (Dark red). The SMS line position is fixed on the *Pion* likelihood curve (to keep the pion misidentification constant), so the dark red region (the region on the kaon likelihood, on the right side of the SMS line) grows bigger as the momentum goes higher and the two likelihoods move closer to each other. But the positive side is that as κ is kept constant while the momentum goes up, we won't get more pions misidentified as kaons. For the LH selector on the other hand, the idea is to keep the Kaon identification efficiency (ε) constant as the momentum changes. In this selector, all tracks on the right hand side of the LH line will fail the Kaon ID. The LH line position is fixed on the *Kaon* likelihood curve (to keep the KID efficiency constant) and this will result in having more pions misidentified as kaons when the momentum goes up, and the two curves overlap more.

In other words, for the SMS κ is kept constant and the KID efficiency drops when the momentum goes higher, and in the LH ε is kept constant and the Pion misidentification increases when the momentum goes higher. Based on the nature of

the analysis, one might find one of these selectors more useful than the other. This behavior is illustrated in Figure 7-8.

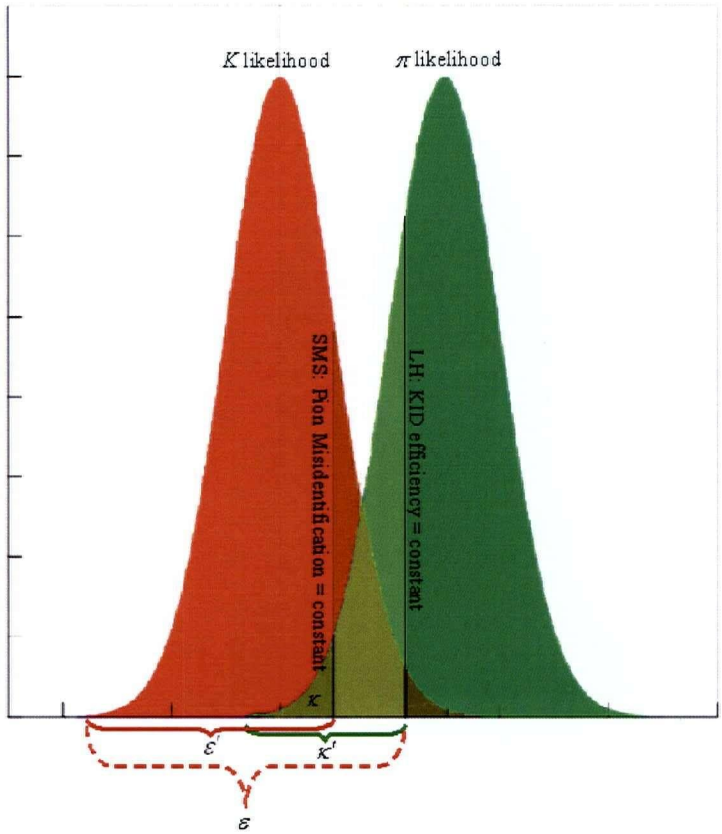

Figure 7-7: ε' is the Kaon identification efficiency for SMS selector and κ is the pion misidentification of the SMS selector. ε is the Kaon identification efficiency for LH selector and κ' is the pion misidentification of the LH selector. The primed ones are the ones that change when the momentum changes.

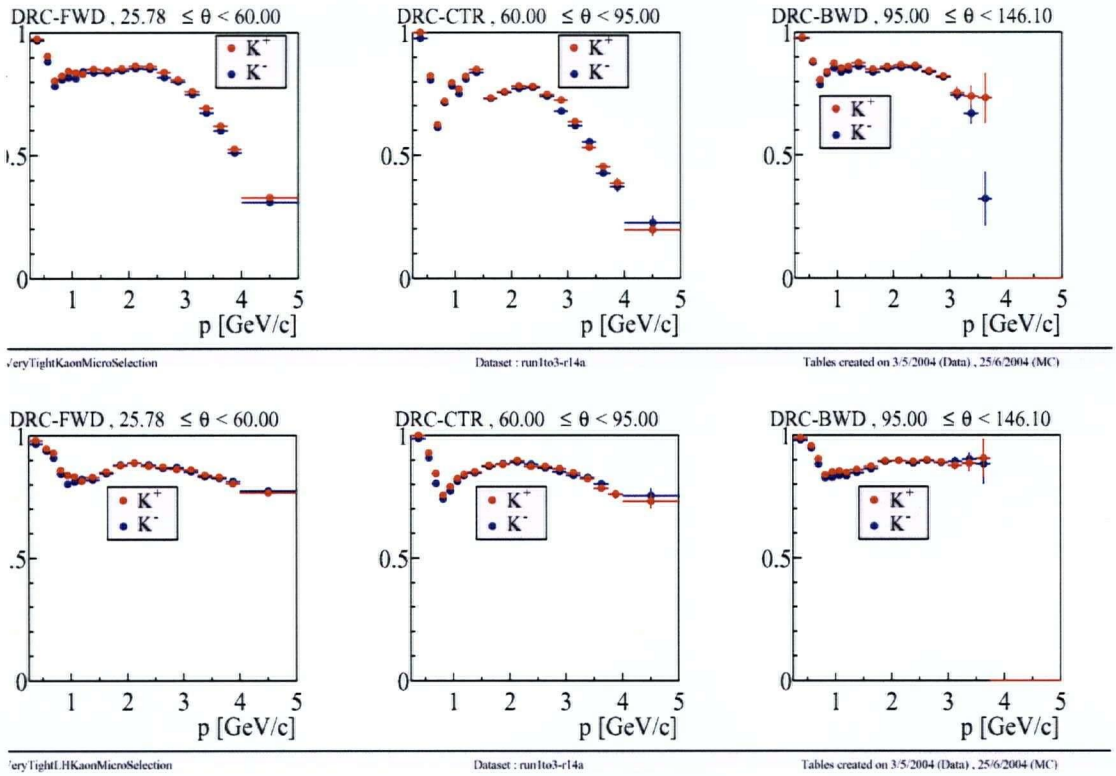

Figure 7-8: Top: Kaon identification efficiency for very tight SMS selector. As you can see the efficiency drops as the momentum increases. Bottom: Kaon identification efficiency for very tight LH selector. The efficiency almost remains constant as the momentum increases

7.5.2 The PIDKaonNNSelector

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.[28] In this particular application for example, the selector learns how to distinguish between different particles using lots of examples that have been fed to it. The algorithm tries to find what variables it can use to make the distinction between particle types, and how those variables are to be used.

As indicated earlier, dE/dx information used by the NN selector is the same as what used by the SMS selector, hence we focus on the main differences. They are the DIRC algorithm used to derive the likelihoods and the way all this information is combined.

The DIRC likelihood is obtained from a global (event) likelihood which is computed at the photomultiplier tube hit level and takes into consideration all the tracks in an event as background for each other, as well as the machine background. It uses the KNET Neural Network to combine all the information. A single continuous variable (close to 1 for identified kaons and to 0 for non-kaons) is the natural output of this algorithm, but for practical reasons, five lists of kaon candidates (similarly to the SMS and LH selectors) are provided to the user.

7.6 Fitting

As described in the next section, the next step we take towards extracting the efficiency of the KID system is to fit the φ peak.

Neglecting interference with nearby resonances[26], the shape of a resonance of nominal mass M_0 and nominal width Γ_0 is described by the Breit-Wigner function. Equation 7-6 describes a functional mass distribution for a resonance mass of M_0 [36]:

$$BW(M) = \frac{M \times M_0 \times \Gamma(M)}{(M_0^2 - M^2)^2 + M_0^2 \Gamma^2(M)}$$

Equation 7-6

Where the resonance width ($\Gamma(M)$) is itself a function of mass,

$$\Gamma(M) = \Gamma_0 \frac{M_0}{M} \left(\frac{q}{q_0} \right)^{2l+1} \times \frac{W_l'^2(q)}{W_l'^2(q_0)}$$

Equation 7-7

Here l is the resonance spin, q is the momentum of its decay products in the resonance rest frame, q_0 is the momentum at $M = M_0$, and $W_l'(q)$ are often neglected Blatt-Weisskopf damping factors.

In this study, a p-wave ($l = 1$) Breit-Wigner $BW(m)$ is used to fit the φ peak:

$$W_1'(q) = \sqrt{\frac{2}{1 + (qR)^2}}$$

Equation 7-8

Where R is the interaction radius of the strong force. Von Hippel and Quigg[37] quote as plausible the interaction radii of 0.25 to 0.75 fm for meson resonance and 0.5 to 1.0 fm for baryon resonance.

---

[26] And Bose-Einstein correlations, which are only important in consideration of the ρ resonance

We have not applied Gaussian smearing to the fit. Gaussian smearing would have very little effect on the final results because the width is dominated by the natural width of the φ, not by detector resolution.

For the background, we have used a parameterized linear function to get an estimation of the number of events. Figure 7-9 shows the fit for the φ peak when we have required the K^- candidate track to pass the very tight SMS selector and no requirement is applied on the identification of the K^+ candidate track. The plot for when the K^+ passes the tight SMS selector and no requirement is applied on K^- is similar.

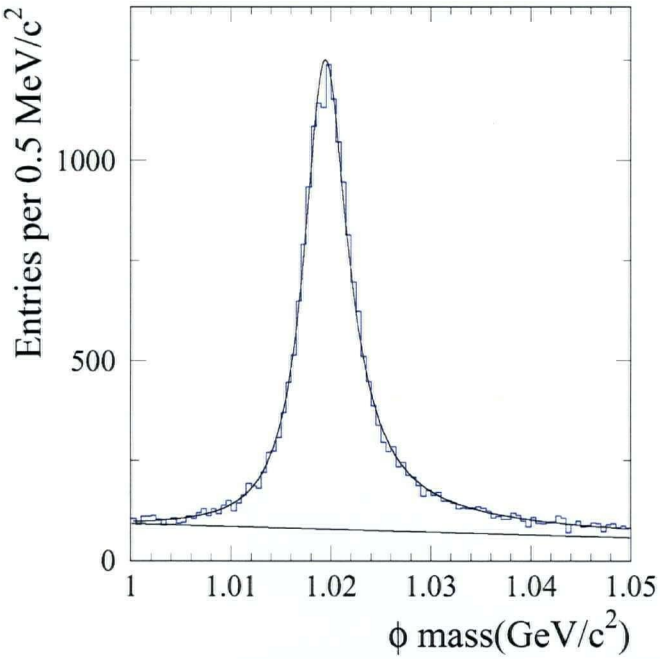

Figure 7-9: Fit results for the φ peak for when the K^- has passed the very tight SMS selector and there is no requirement on the K^+ track identification. The solid line indicates the background.

7.7 Extracting the Efficiencies

After the fitting, we extract the number of φs in the peak by adding up the value of the fit function at each bin minus the value of the background function at that bin.

The efficiency of each selector for positively charged tracks is then calculated by dividing the number of φs in each histogram applying different selection criteria on the K^+, by the total number of φs in the histogram that applies no requirement on the K^+. The same method is applied to determine the efficiency for negatively charged kaons.

$$\varepsilon = \frac{\#\varphi,\ \text{VT - SMS on K}^-}{\#\varphi, \text{no requirement on K}^-}$$

Equation 7-9

As we mentioned in Section 7.1 we do this for 4 different situations:

- To study the KID system efficiency in identifying positively charged kaons (K^+) we require the negatively charged kaon track (K^-) to pass the very tight SMS selection (this is part of the method for isolating the signal as mentioned in 5.4) and then we require the K^+ track to pass different modes of different selectors and extract the efficiency at each step. (Figure 7-9-Figure 7-12.)
- To study the KID system efficiency in identifying negatively charged kaons (K^-) we do exactly the same thing; except that we switch the places of positively charged kaons and negatively charged ones. The plots for this case are similar.
- We then redo both the steps above, requiring the tracks to make it to the DIRC.

The extracted efficiencies are summarized in tables 7-2 and 7-3.

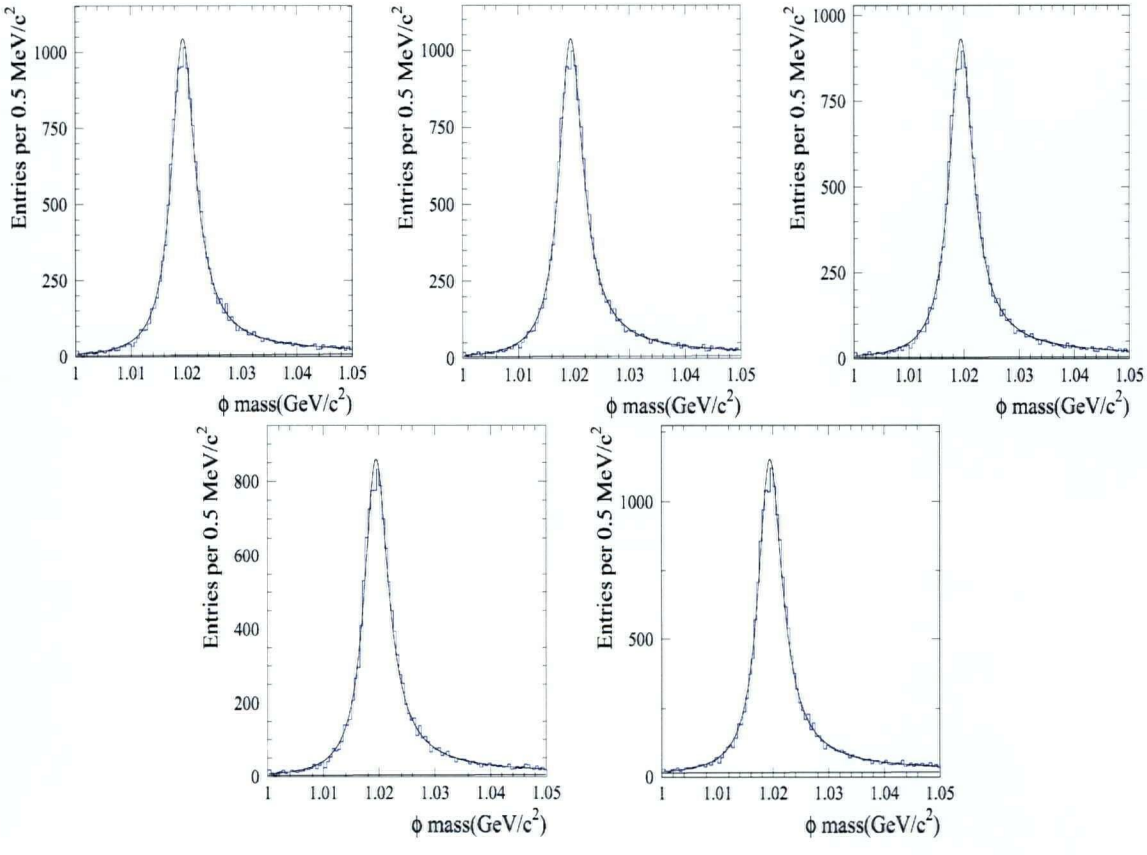

Figure 7-10: The φ peak for when the K^- has passed the very tight SMS selector. K^+ passes: from top left: the very loose SMS selector, loose SMS selector, tight SMS selector, very tight SMS selector, not a Pion SMS selector

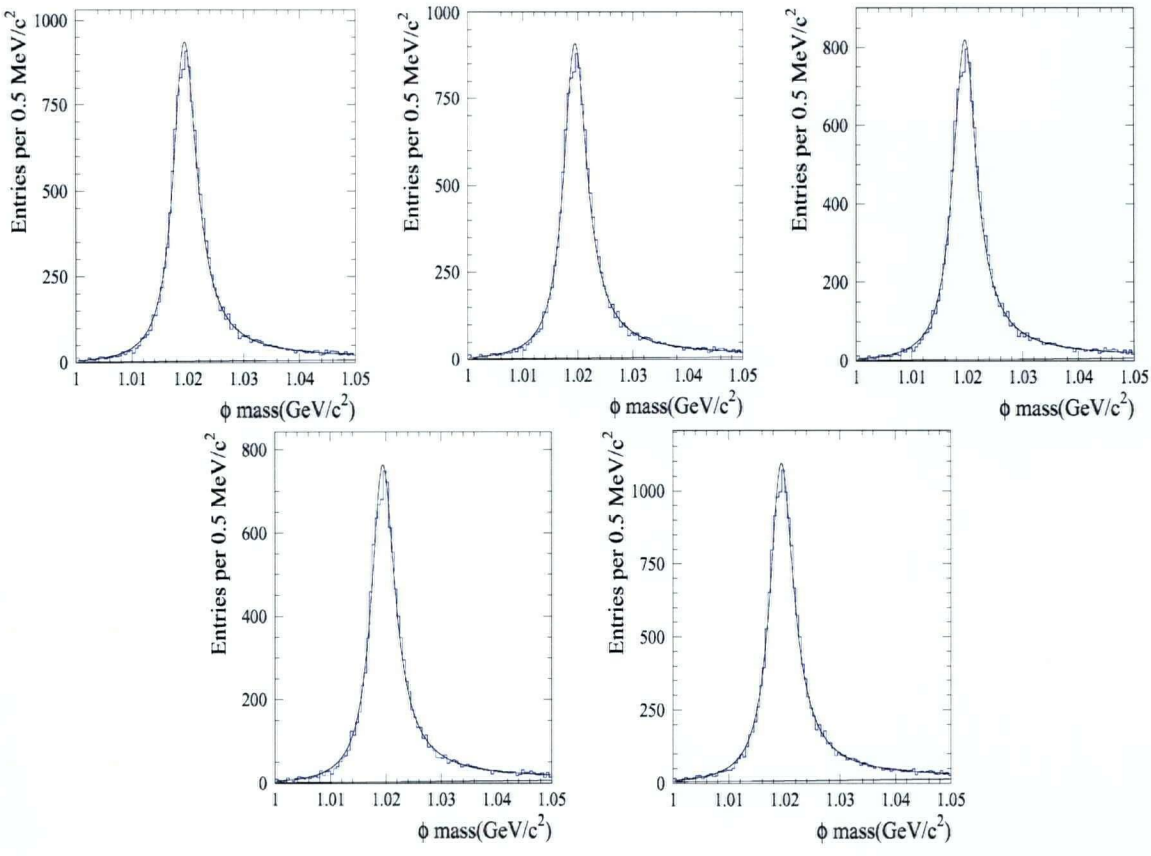

Figure 7-11: The φ peak for when the K^- has passed the very tight SMS selector. K^+ passes: from top left: the very loose NN selector, loose NN selector, tight NN selector, very tight NN selector, not a Pion NN selector

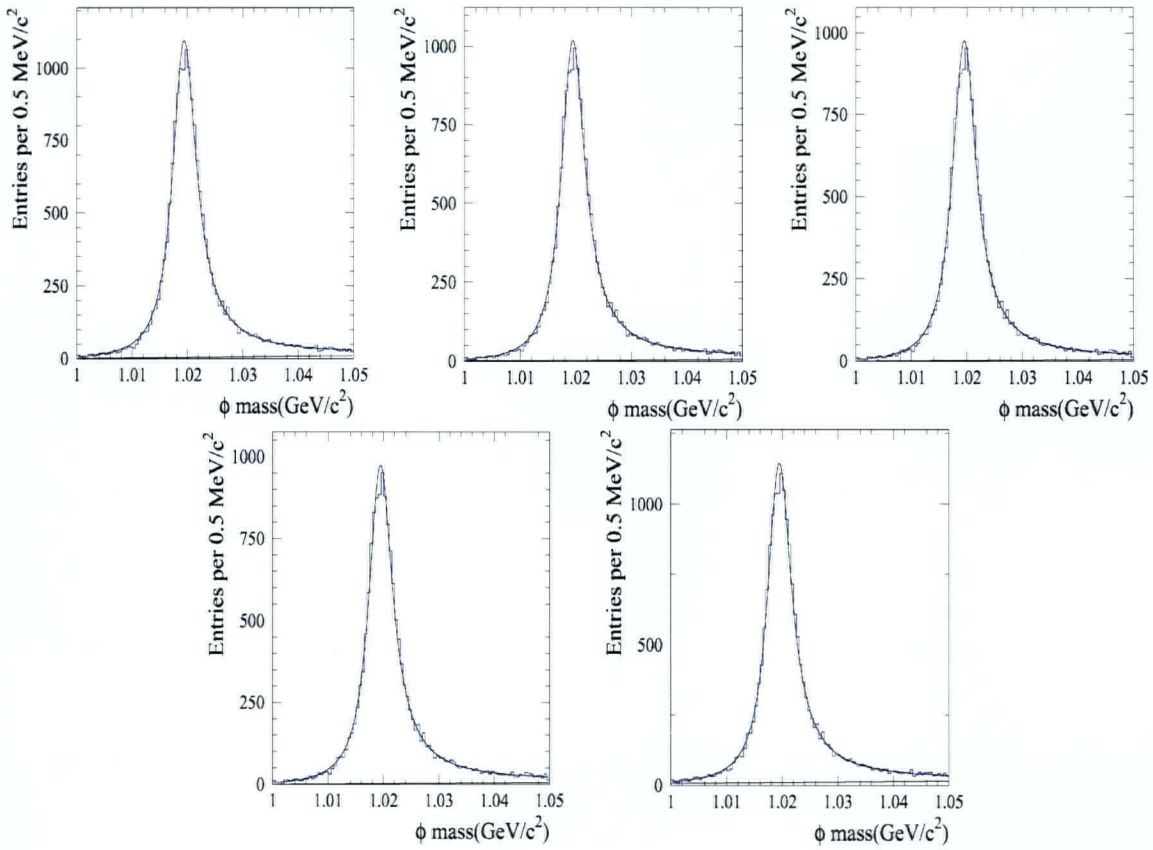

Figure 7-12: The φ peak for when the K^- has passed the very tight SMS selector. K^+ passes: from top left: the very loose LH selector, loose LH selector, tight LH selector, very tight LH selector, not a Pion LH selector

		$\varepsilon_{K^+,NoD}$	$\varepsilon_{K^+,D}$	$\varepsilon_{K^-,NoD}$	$\varepsilon_{K^-,D}$
SMS	Very loose	86.7%	88.4%	89.3%	90.6%
	Loose	86.4%	87.0%	88.3%	88.7%
	Tight	77.9%	80.3%	80.0%	81.7%
	Very tight	71.9%	74.1%	74.0%	75.7%
	Not a pion	96.3%	96.2%	96.7%	96.6%
NN	Very loose	76.9%	77.7%	83.3%	83.8%
	Loose	74.5%	75.5%	81.3%	81.9%
	Tight	67.6%	69.0%	75.0%	76.0%
	Very tight	63.2%	64.8%	70.8%	72.0%
	Not a pion	90.9%	91.3%	94.6%	94.9%
LH	Very loose	96.1%	96.4%	97.2%	97.4%
	Loose	91.7%	92.3%	93.5%	93.8%
	Tight	85.4%	86.6%	87.8%	88.5%
	Very tight	82.0%	83.3%	84.4%	85.2%
	Not a pion	81.8%	83.1%	84.1%	85.0%

Table 7-2: Data efficiencies for five different modes of the three selectors.

		$\varepsilon_{K^+,NoD}$	$\varepsilon_{K^+,D}$	$\varepsilon_{K^-,NoD}$	$\varepsilon_{K^-,D}$
SMS	Very loose	88.4%	89.6%	90.5%	91.5%
	Loose	87.1%	87.1%	89.6%	89.6%
	Tight	79.1%	80.7%	81.3%	82.6%
	Very tight	73.3%	74.8%	75.4%	76.5%
	Not a pion	95.5%	95.4%	96.0%	96.0%
NN	Very loose	81.6%	81.7%	87.9%	88.0%
	Loose	79.5%	79.7%	86.0%	86.2%
	Tight	73.0%	73.9%	80.7%	81.2%
	Very tight	68.6%	69.7%	77.0%	77.8%
	Not a pion	92.9%	93.0%	96.1%	96.2%
LH	Very loose	96.5%	96.6%	97.5%	97.5%
	Loose	92.6%	92.7%	94.5%	94.6%
	Tight	86.9%	87.8%	89.9%	90.0%
	Very tight	82.6%	83.1%	86.1%	86.4%
	Not a pion	81.9%	82.6%	87.7%	86.1%

Table 7-3: MC efficiencies for five different modes of the three selectors.

As you can see the best performance is seen in the LH selector. One would suspect that the NN selector should have performed the best, since it seems to be the most intelligent way of approaching the issue. But the problem is that because the examples that have been given to the network might have had some sort of biases in them or just the fact that the examples didn't completely cover the vast variety of possibilities that could happen in the detector, the network might develop a sensitivity towards some variables that are not really the key in distinguishing between different particles. This will make the selector falsely reject or accept particles as Kaons. As for the SMS selector, it is designed to keep the Pion misidentification efficiency κ constant (7.5.1), and this will result in lower Kaon identification efficiency.

One can also see that the best agreement between Data and MC is achieved in the LH selector while the NN selector is the worst in this respect. The efficiency of the NN selector is highly over-estimated in MC, which also indicates that the neural network algorithm isn't educated enough about what *really* happens in the detector.

Table 7-4 provides the ratio of the efficiencies (data/MC) in Table 7-2, Table 7-3. The overall average of efficiency ratios to the second decimal number for the SMS selector is 0.99, for the NN selector is 0.95 and for the LH selector is 0.99 which

suggests that MC best represents SMS and LH selector performances. We have to note that although the SMS and LH have equal data/MC average efficiency ratios, if we were to neglect the $\varepsilon_{K^+,D}$ in *LH- not- pion* selector, we would have the best agreement between data and MC in LH selector.

		$\varepsilon_{K^+,NoD}$	$\varepsilon_{K^+,D}$	$\varepsilon_{K^-,NoD}$	$\varepsilon_{K^-,D}$
SMS	Very loose	0.98	0.99	0.99	0.99
	Loose	0.99	1.00	0.99	0.99
	Tight	0.98	1.00	0.98	0.99
	Very tight	0.98	0.99	0.98	0.99
	Not a pion	1.01	1.01	1.01	1.01
NN	Very loose	0.94	0.95	0.95	0.95
	Loose	0.94	0.95	0.95	0.95
	Tight	0.93	0.93	0.93	0.94
	Very tight	0.92	0.93	0.92	0.93
	Not a pion	0.98	0.98	0.98	0.99
LH	Very loose	1.00	1.00	1.00	1.00
	Loose	0.99	1.00	0.99	0.99
	Tight	0.98	0.99	0.98	0.98
	Very tight	0.99	1.00	0.98	0.99
	Not a pion	1.00	1.00	0.96	0.99

Table 7-4: The data/MC ratio of the efficiencies.

Figure 7-13 and Figure 7-14 give a better sense of these behaviors and also provide an estimate of the Background to Signal ratio for each selector. We calculate the Background to Signal ratio by dividing the sum of background in all bins by the total number of φ mesons.

As can be seen, the LH selector has the smallest Background/Signal ratio while having the highest efficiency. We can also see that the MC underestimates the background in the events compared to data.

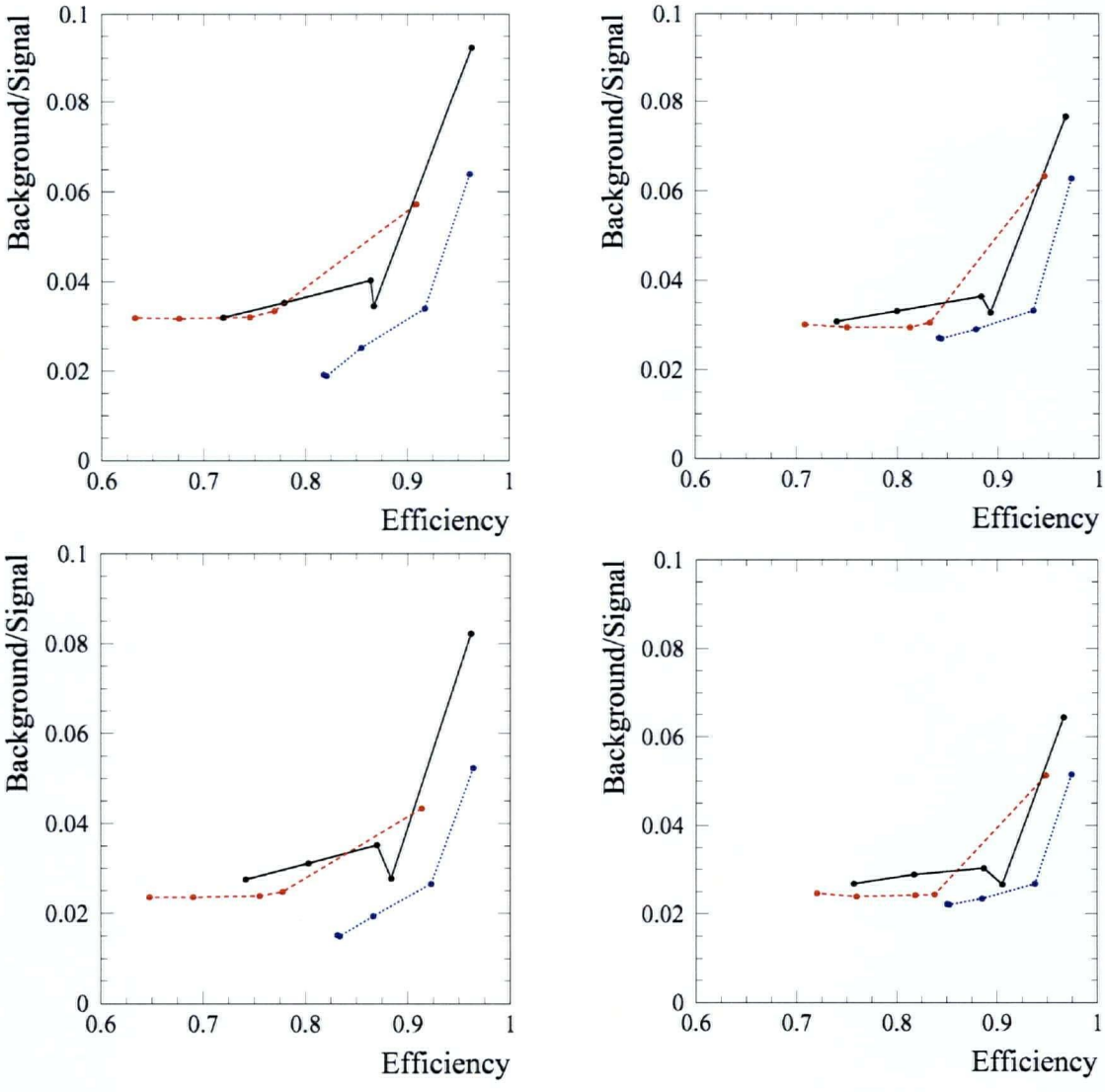

Figure 7-13: A comparison between behaviors of different selectors in Data. The black (solid) line is SMS, the red (dashed) one is NN and the blue (dotted) line is the LH selector. From top left the efficiencies are: $\varepsilon_{K^+,NoD}$, $\varepsilon_{K^+,D}$, $\varepsilon_{K^-,NoD}$ **and** $\varepsilon_{K^-,D}$.

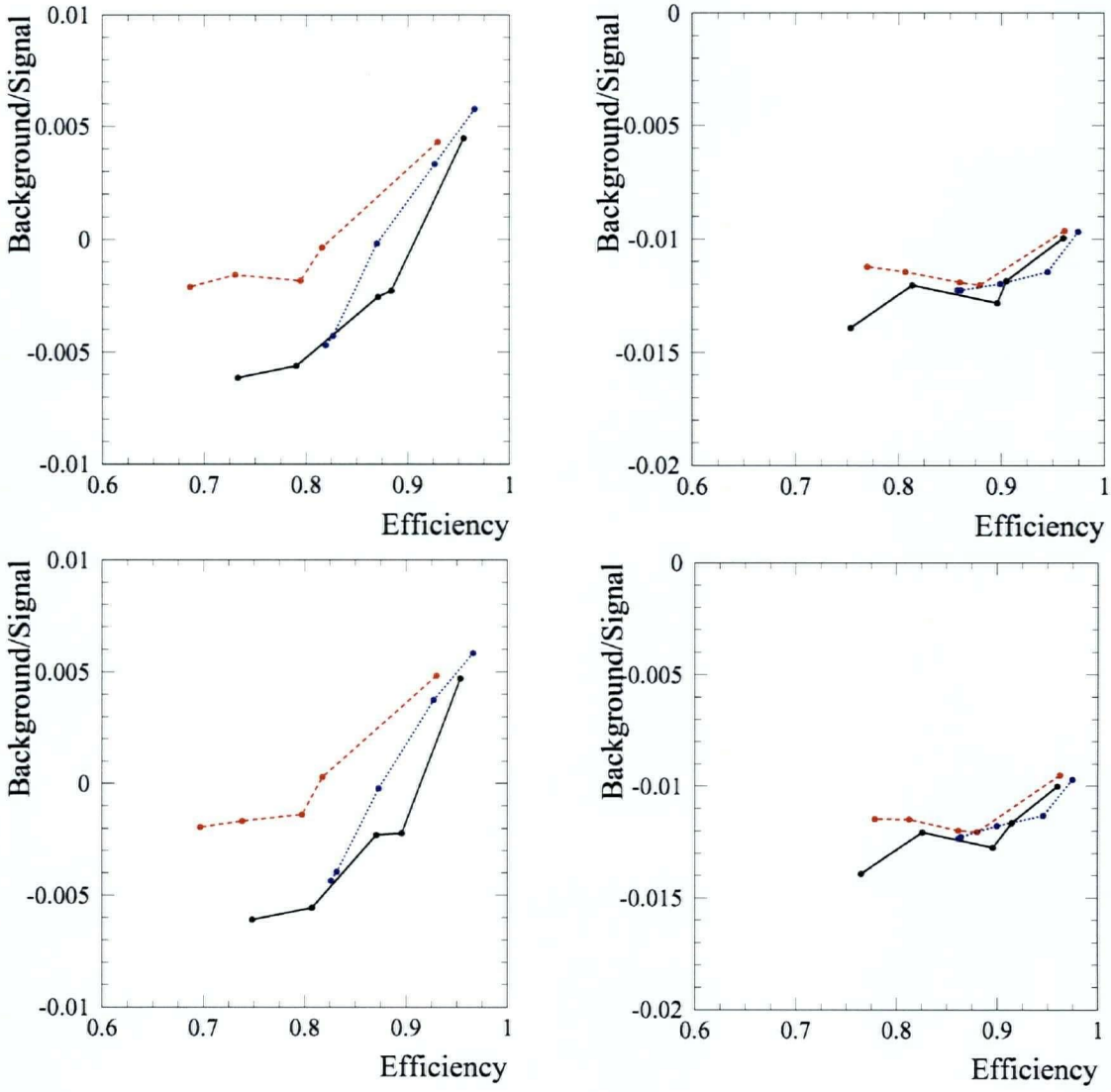

Figure 7-14: A comparison between behaviors of different selectors in MC. The black (solid) line is SMS, the red (dashed) one is NN and the blue (dotted) line is the LH selector. From top left the efficiencies are: $\varepsilon_{K^+,NoD}$, $\varepsilon_{K^+,D}$, $\varepsilon_{K^-,NoD}$ **and** $\varepsilon_{K^-,D}$.

8

Summation

This thesis uses the control sample $e^+e^- \to \varphi\gamma, \varphi \to K^+K^-$ to study some detector behaviors of the BaBar detector and is split into two distinct, yet equally important portions. The first half is dedicated to studying the extra energies in the event. Both the shape and the magnitude of these energies have been studied in detail in Chapter 6. A comparison was also made between data and Monte Carlo (MC) to study the quality of the simulation. The following highlights the results observed in this work:

- The extra neutral energy is poorly simulated in both the shape and quantity near a charged track. This is due to weak simulation of hadronic showers of a charged track in the EMC.
- The extra neutral energy, extra charged energy and number of extra charged tracks are underestimated in MC.
- About 90% of events have no extra charged tracks in them. The Data/MC agreement of total extra energy in these events is much better than that of the events with extra charged tracks due to the absence of extra charge energy disagreement.
- A Data/MC disagreement was observed near the high energy photon which is most likely due to the poor simulation of electromagnetic showers of the photon.
- Applying the cuts mentioned in BAD #633 to choose "good" tracks and "good" clusters worsened the Data/MC agreement in all cases.

The aim of the second half of this analysis is to extract the efficiency of Kaon identification (KID) system used in BaBar experiment. The KID system has three families of selectors, each using different algorithms to identify the "kaonness" of a track. These families and their modes are fully described in Chapter 7. Here is the summary of the results observed in this study:

- The best efficiency is seen in *PidKaonLHSelector* which is designed to keep the KID efficiency constant as the momentum changes.
- The worst performance is seen in *PidKaonNNSelector*. This implies that the network is not very well trained and most likely has developed some false

sensitivity toward some not important variables and/or missing sensitivity towards some crucial ones.

Bibliography

[1] T.T. Wu and C.N. Yang, "Phenomenological Analysis of Violation of CP Invariance in Decay of K^0 and $\overline{K}^0$", Phys. Rev. Lett. **13**, 380 (1964).

[2] J.H Christenson, J.W. Cronin, V.L. Fitch, and R. Turlay, Phys. Rev. Lett. **13**,138 (1964).

[3] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", Phys. Rev. Lett. **10**, 531 (1963).

[4] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interactions", Prog. Th. Phys. **49**, 652 (1973).

[5] D. Boutigny *et al.* [BABAR Collaboration], " BABAR Technical Design Report", SLAC-R-0457.

[6] B. Aubert *et al.* [BABAR Collaboration], "The BABAR detector", Nucl. Instrum. Meth. A **479**, 1 (2002).

[7] J. Seeman *et al.* "Status report on PEP-II performance," *Presented at 7th European Particle Accelerator Conference (EPAC 2002), Vienna, Austria, 26-30 Jun 2000.* SLAC-PUB-8629.

[8] R.Claus, P. Grosso, R. T. Hamilton, M. E. Huffer, C. P. O'Grady, J. J. Russell and I. Scott, SLAC-PUB-8134 *Presented at 11th IEEE NPSS Real Time Conference (Santa Fe 99), Santa Fe, NM, 14-18 Jun 1999.*

[9] A. Ryd *et al.* [BABAR Computing Group Collaboration], eConf **C0303241**, MODT012 (2003) [arXiv:cs.dc/0306069].
T. Glanzman, J. Bartelt, R. F. Cowan, S Dasu, G. Gordidier, F. di Lodvico and F. Safai Tehrani (BABAR Computing Group Collaboration), SLAC-PUB-8388.

[10] A. Ryd *et al.* "Trigger and filter documentation for the Run1 data." BAD 194, Apr. 2004.

[11] B. Jacobson [BABAR Collaboration], Nucl. Instrum. Meth. A **389**, 1 (1997)

[12] D. H. Wright *et al.* [BABAR Computation Group Collaboration], eConf **C0303241**, TUMT0006 (2003) [arXiv:hep-ph/0305240].

[13] BABAR PID Group, http://www.slac.stanford.edu/BFROOT/www/Physics/Tools/Pid/pid.html

[14] http://paw.web.cern.ch/paw/

[15] http://root.cern.ch/root/

[16] S.L. Glashow, Nucl. Phys. **22**, 579 (1961); Weinberg, Phys. Rev. Lett. 19, 1264 (1967); A. Salam, "Proceedings of the 8th Nobel symposium", ed. N. Swartholm, Almquist and Wiksells, Stockholm (1968).

[17] K. Hagiwara *et al.* [Particle Data Group Collaboration], Phys. Rev. D **66**, 010001 (2002).

[18] "Proceedings of the First Workshop on the CKM Unitarity Triangle", CERN-2003-002, Apr 2003. To appear as CERN Yellow Report. Editors – M. Battaglia, A.J. Buras, P. Gambino, A.Stocchi. (hep-ph/034132).
[19] L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983).
[20] "The BABAR physics book", P. H. Harrison and H. R. Quinn, eds., SLAC-R-504 (1998), and refrences therein.
[21] "Supporting Document for the Summer 2004 $Sin2\beta$ Analysis", BABAR Analysis Document **929** v.9(2004).
[22] N. Kuznetsova, J. Richman, A. Ryd, "Search for the Rare decays $B \to Kl^+l^-$ and $K^*l^+l^-$ ", BABAR Analysis Document **270**.
BABAR Collaboration, B. Aubert et al., Phys. Rev. Lett. **91**, 221802 (2003).
[23] P. Jackson, B. Kowalewski [BABAR Collaboration], "Search for $B^+ \to K^+\nu\bar{\nu}$ using semi-leptonic tags", BABAR Analysis Document **293** v.11(2004).
[24] Y. Grossman, Z. Ligeti and E. Nardi, "First limit on inclusive $B \to X_s\nu\bar{\nu}$ decay and constraints on new physics" hep-ph/9510378.
[25] G. Buchalla, G. Hiller and G. Isidori,"Phenomenology of non-standard Z couplings in exclusive semileptonic $b \to s$ transitions" hep-ph/0006136 v2.
[26] S. Spanier and G. Mancinelli, "Kaon Identification in the BABAR experiment", BABAR Analysis Document **116**.
[27] http://rd11.web.cern.ch/RD11/rkb/PH14pp/node62.html#61
[28] Christos Stergiou and Dimitris Siganos , "Neural Networks", http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.htm
[29] Doug Thiessen [BABAR Collaboration], "A new MuMuGamma Control Sample", BABAR Analysis Document **268** v.1 (2001).
[30] F. Fabozzi *et al.* [BABAR Collaboration], "Muon Identification in the BABAR Experiment," BABAR Analysis Document **60** v.01 (2000).
[31] Dominique Fortin *et al.* [BABAR Collaboration], "Particle selection criteria, optimized for visible energy measurement", BABAR Analysis Document **633** v.4 (2003).
[32] Donald H. Perkins "Introduction to High Energy Physics", 4th edition.
[33] J. Albert *et al.*, "BaBar Kinkfinding Software", BABAR Analysis Document **31**.
[34] S. Yellin, "Using Global Likelihood to Compare DIRC PID hypotheses", DIRC Note **127**.
[35] T. Haas *et al.*, "Production of $\pi^\pm$, $K^\pm$ and $p/\bar{p}$ at sqrt(s) = 10.54 GeV and in Hadronic $\Upsilon(4S)$ Decays", BABAR Analysis Document **158**.
[36] Alexander V. Telnov, Andrei V. Gritson, "Branching Fraction in $B \to \varphi h$ and Search for Direct CP Violation in $B^\pm \to \varphi K^\pm$," BABAR Analysis Document **406** v.11(2003).
[37] F. Von Hippel and C. Quigg, "Centrifugal-Barrier Effects In Resonance Partial Decay Widths, Shapes, and Product Amplitudes", Phys. Rev. D **5**, 624(1972).