A Comparison of Multi-variate Probability Density Estimate Methods in Neutral Pion Discrimination

M. Sc Thesis
in
Engineering Physics
University of Gaziantep

Supervisor
Assist. Prof. Dr. Ayda BEDDALL

By
Zekeriya UYSAL
December 2006

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Sadettin ÖZYAZICI
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Zihni ÖZTÜRK
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Ayda BEDDALL
Major Supervisor

Examining Committee Members:

Prof. Dr. Zihni ÖZTÜRK _____

Assoc. Prof. Dr. Nihat YILDIRIM _____

Assist. Prof. Dr. Andrew BEDDALL _____

Assist. Prof. Dr. Ayda BEDDALL _____

Assist. Prof. Dr. Bahattin KANBER _____

Abstract

A Comparison of Multi-variate Probability Density Estimate Methods in Neutral Pion Discrimination

UYSAL, Zekeriya

M. Sc in Engineering Physics

Supervisor: Assist. Prof. Dr. Ayda BEDDALL

December 2006, 50 pages

A number of multi-variate PDE (Probability Density Estimator)methods are compared for the discrimination of signal from background in the selection of neutral pion candidates reconstructed at the ALEPH experiment at CERN. In this case-study, the question "which method is the best choice?" reveals that the answer depends strongly on the size of the data set used to train and optimise the method, and the required simplicity of the algorithm.

Key words: Multi-variate Probability Density Estimators (PDE), Pion discrimination.

Özet

Yüksüz Pion Adaylarının Ayıklanmasında Çok Değişkenli ihtimal Yoğunluğu Tahmini Yöntemlerinin Karşılaştırılması

UYSAL, Zekeriya
Yüksek Lisans Tezi, Fizik Mühendisliği
Tez Yöneticisi: Y. Doç. Dr. Ayda BEDDALL
Aralık 2006, 50 sayfa

Çok değişkenli ihtimal yoğunluğu tahmini yöntemleri, CERN deki ALEPH deneylerinde elde edilen verilerdeki yüksüz pion adaylarının diğer sinyallerden (background) ayıklanmasında kullanılmış ve karşılaştırılmıştır. Bu çalışmada, "hangi metod en iyi seçimdir?" sorusu, metodu optimize ederken kullanılan veri kümesinin boyutuna ve algoritmanın basitliğine bağlı olarak değişkenlik gösterir.

Anahtar kelimeler: Çok Değişkenli ihtimal Yoğunluğu Tahmini Yöntemleri (PDE), Piyon Ayıklanması

Acknowledgements

I would like to thank Assist. Prof. Dr. Ayda Beddall and Assist. Prof. Dr. Andrew Beddall for their support and guidance during the preparation of this thesis.

Also special thanks to Instructor Ahmet Bingül and Res. Assist. Yılmaz Durmaz for their helps.

And very special thanks to my Export Department Manager Mrs. Fitnat Tekerekoğlu for her understanding and spiritual support.

Finally i would like to very special thanks to my family and my girlfriend Arzu Batur for their spiritual support

Table of Contents

Chapter

1 Introduction 1

2 Particle Detectors 2
2.1 Accelerators . . . 2
2.2 To See a Particle . . . 3
2.3 The Passage of Charged Particles Through Matter . . . 3
2.3.1 Energy Loss for Heavy Particles . . . 3
2.3.2 Energy Loss for Electrons . . . 4
2.3.3 Multiple Coulomb Scattering . . . 6
2.3.4 Čerenkov Radiation . . . 6
2.4 The interaction of High-Energy Photons with Matter . . . 6
2.5 Particle Detectors . . . 7
2.5.1 Electroscope . . . 8
2.5.2 Scintillation Counter and Photomultiplier . . . 8
2.5.3 Geiger-Müller Tube . . . 8
2.5.4 Proportional Counters . . . 10
2.5.5 Drift Chambers . . . 10
2.5.6 Ionisation Chambers . . . 11
2.5.7 Bubble Chambers . . . 11
2.5.8 Electromagnetic Shower Detectors and Calorimeters . . . 11

3 The ALEPH Detector 14
3.1 Overview . . . 14
3.2 The Central Tracking Chambers . . . 16
3.3 The Vertex Detector (VDET) . . . 16
3.4 The Inner Tracking Chamber (ITC) . . . 16
3.5 The Time Projection Chamber (TPC) . . . 16
3.6 The Electromagnetic Calorimeter (ECAL) . . . 17
3.7 The Hadron Calorimeter (HCAL) and Muon Chambers . . . 19
3.8 The Superconducting Solenoid . . . 20
3.9 The Luminosity Monitors . . . 20
3.10 The Silicon Calorimeter . . . 20
3.11 The Luminosity Calorimeter . . . 21
3.12 BCAL . . . 21

4 Event Reconstruction and Simulation 22
4.1 Introduction . . . 22
4.2 Track Reconstruction . . . 22

4.2.1 TPC Coordinate Determination . . . 24
4.2.2 Coordinates . . . 27
4.2.3 TPC Track Finding and Fitting . . . 27
4.2.4 TPC-ITC-VDET Track Association . . . 28
4.2.5 Ionisation Energy Loss . . . 28
4.3 Calorimeter Reconstruction . . . 29
4.3.1 Calorimeter Clusters and Objects . . . 29
4.4 γ/π^0Reconstruction and Identification . . . 29
4.4.1 G-clusters . . . 30
4.4.2 Photon Position Calculation . . . 30
4.4.3 Photon Energy Calculation . . . 31
4.4.4 π^0Reconstruction . . . 31
4.4.5 Mass Constraint . . . 31
4.4.6 Merged Photons . . . 32
4.5 Event Simulation . . . 32
4.5.1 KINGAL . . . 32
4.5.2 GALEPH . . . 32

5 Event Selection Using Adaptive Gaussian Kernels 33
5.1 Introduction . . . 33
5.2 Kernel Estimation . . . 33
5.2.1 Fixed Kernel Estimation . . . 34
5.3 Multivariate Kernel Estimation . . . 35
5.3.1 Fixed Kernel Estimation . . . 35
5.3.2 Adaptive Probability Density Estimation . . . 35

6 Comparison of PDE Methods 37
6.1 PDE Methods . . . 37
6.1.1 The Histogrammed PDE, HPDE . . . 38
6.1.2 The Smoothed Histogrammed PDE, SHPDE . . . 38
6.1.3 The Kernel PDE, KPDE . . . 39
6.1.4 The Histogrammed Kernel PDE, HKPDE . . . 39
6.1.5 The Range-Search PDE, RSPDE . . . 39
6.2 PDE Performances . . . 40
6.2.1 $\varepsilon \times \mathcal{P}$ Performance . . . 43
6.2.2 Timing . . . 45

7 Summary and Conclusion 47

References 48

Appendix

A π^0Reconstruction from Photon Pairs γ_1,γ_2 49

B χ^2 and Lagrange Multiplicator 50
B.1 Method of Lagrange Multipliers for Maxima and Minima 50

List of Figures

Figure

2.1 The variation of intensity of a beam of particles with depth 5
2.2 The variation in the rate of energy loss for a charged particle with depth 5
2.3 Schematic of a photomultiplier tube 9
2.4 Schematic of proportional and drift chamber 12

3.1 Schematic view of the ALEPH detector 15
3.2 Energy resolution of the ECAL. 19

4.1 The flow of data in the ALEPH experiment. 23
4.2 Reconstruction of tracks from coordinates in the TPC. 25
4.3 A TPC pad row cluster formed by two nearby tracks. 26
4.4 TPC pad row subclusters formed by two nearby tracks. 26
4.5 Helix parameters used in the TPC tracking algorithms. 28
4.6 Energy deposits of photons in the ECAL. 30
4.7 Mass constraint χ^2. 31

6.1 Flowchart of the PDE programs 41
6.2 Parameters used in the discrimination between signal and background. 42
6.3 Comparison of optimal $\varepsilon \times \mathcal{P}$ values 44
6.4 Comparison of average run-times 46

List of Symbols

e^- : Electron

e^+ : Positron

π^0 : Neutral pi-meson

γ : Photon

θ_{12} : Opening angle of the photons

σ : Gaussian width

β : A constant that equals v/c

Z : Atomic mass of the nucleus

N_0 : Avagadro's number

X_0 : Radiation length, g/cm^2

c : Speed of light

n : Index of refraction

n : Number of photons or pions

w_1 : Measured energy of the first photon

w_2 : Measured energy of the second photon

χ^2 : Chi-squre distribution

E_1 : Energy of first photon which is searched

E_2 : Energy of second photon which is searched

λ : Lagrange multiplicator

Chapter 1

Introduction

Multi-variate probability density estimators (PDEs) have become important methods in the effective discrimination of signal and background in high energy physics data analysis. Several alternative multi-variate PDE methods have appeared in the literature in recent years, demonstrations of these methods often include a comparison with results from a neural network as neural networks generally give the best performance. A direct comparison between some of these PDE methods, using data generated from a toy Monte-Carlo, can be found in Reference [1].

For the main study presented in this thesis, a number of PDE methods are employed for the three-dimensional discrimination of signal from background in the selection of neutral pion candidates reconstructed at the ALEPH experiment at CERN using a full physics and detector simulation. This study, therefore, provides a real-life case-study with data that contains the complexities of particle dynamics, and the correlations that are often present between variables in high energy physics. The study aims to provide the reader with some useful experience and discussion to aid the researcher in the choice of a PDE method. The issues covered are: algorithm simplicity, discrimination performance, importance of statistics, and the time it takes to train and optimise the algorithm (run-time).

Before presenting the main study of this thesis, some introductory information is given. In Chapter 2, a general overview of particle detectors is given. Chapter 3 describes in detail the ALEPH detector. In Chapter 4, event reconstruction is described. In Chapter 5, an Adaptive Multi-variate Gaussian Kernel PDE method is described, this is an example of a contemporary PDE method used in high energy physics. A comparison of the performance of a number of PDE methods is given in Chapter 6. Finally, the conclusion of the study is given in Chapter 7.

Chapter 2

Particle Detectors

2.1 Accelerators

All accelerators employ electric fields and/or magnetic field to accelerate stable charged particles (electrons, protons or heavier ions) to high energy. Many types have been developed for the study of nuclear reactions. Rather than giving a charge one large acceleration with high voltage, it can be brought to high speed by succession of accelerations through relatively small potential differences, as in the *linear accelerators*. All modern accelerators are circular, *cyclic accelerator*, or nearly so. The particles are constrained in a vacuum pipe bent into a torus which threads a series of electromagnets, providing a field normal to the plane of the orbit. There are a variety of significant reasons for building accelerators with the highest practical beam energies. Some of these are:

(1) At high energies the shorter de Broglie wavelength h/p of the beam particles allows the possibility of probing the structure of target particles in increasingly fine detail.

(2) Since there is as yet no general theory for strong interactions and only an incomplete one for the weak interactions, the high-energy behavior of cross sections, etc., can only be determined experimentally. Moreover there are reasons to hope that in the limit of very high energies, the strong interactions will be simpler to treat theoretically, which may allow an important breakthrough in their understanding. In any case it does not seem likely that the strong interactions will ever be understood completely until we can experimentally study their asymptotic behavior at very high energies.

(3) The production of secondary particles (pions, kaons, antileptons, neutrinos, etc.) is generally more copious for higher beam energies. This facilities the study of their interactions and allows them to be studied

at higher energies. It also becomes practical to make beams of short-lived particles such as hyperons at high energies, and thus make detailed studies of their interactions possible.

(4) Cross section for weak interactions generally increase with energy. This makes experiments to study neutrino interactions, for example, considerably easier at high energies.

2.2 To See a Particle

In the 1930s the physicist perceived the atom as being composed of particles of three kinds: electrons, protons and neutrons, each characterized by its own mass and electric charge. The development of suitable instrumentation plays a principal role in the study of sub-atomic particles. Not even the most powerful microscope will enable us to see sub-atomic particles. But there are certain instruments which so unmistakable reveal the presence of these particles and trace their paths that we can indeed claim to be able to *see* them. Here we shall briefly summarize the important aspects of the behavior of charged particles and high-energy photons in matter to provide a basis for the material on detectors to be discussed later.

2.3 The Passage of Charged Particles Through Matter

In addition to nuclear interactions, charged particles passing through matter lose energy through a variety of electromagnetic processes. For all charged particles except electrons, collision with electrons are the most important process. (Electrons behave quite differently than the heavier particles because of their small mass and the absence of strong interactions.)

2.3.1 Energy Loss for Heavy Particles

In an encounter with an atomic electron a heavy, charged particle can excite it to a higher level or remove it completely form the atom. The trail of free electrons and ions thus formed by charged particle forms the basis for most techniques for detecting such particles (e.g., bubble chambers, scintillation counters, spark chambers, photographic emulsions). Occasionally the electrons receive enough energy to make them visible as distinct tracks in bubble chambers

and cloud chambers. These electrons are called *delta rays*. The energy loss due to ionisation can be calculated in a relatively straightforward fashion.

It is found that the rate of energy loss, dE/dx, of a particle is a function only of its velocity and its charge squared. The typical behavior of a beam of charged particles (initially monochromatic) in matter is shown in Figure 2.1. For hadrons there is a continuous attenuation due to nuclear interactions in the medium; this does not occur with muon beams. On average the particles have a rather well-defined range R_0, but statistical fluctuations in the energy loss cause some variation about R_0. This effect is called *range straggling*; r_{St} is a measure of the fluctuation in range.

The behavior of dE/dx, the rate of energy loss, is shown Figure 2.2. For fast particles ($v \approx c$), the rate of loss is almost constant and rather small. As the practice slows down, however, $|dE/dx|$ rapidly increases. For $v \approx c$ in hydrogen $dE/dx \approx 4.1\ MeV/g * cm^{-2}$, for most other substances $-dE/dx \approx 2\ MeV/g * cm^{-2}$, at $R \to R_0$.

2.3.2 Energy Loss for Electrons

Because of their small mass, electrons lose a significant fraction of their energy by radiation or *bremsstrahlung* (breaking radiation). Bremsstrahlung occurs when a charged particle is accelerated in the electric field of a nucleus. This effect is more important in high-Z materials. The average rate of energy loss due to radiation goes approximately as Z^2/m^2, where m is the rest mass of incident particle, and Z the atomic number of nucleus. Thus radiation loss is much less important for the heavier particles.

On average a high-energy electron traversing $x\ g/cm^2$ of material will lose a fraction of its energy

$$\frac{dE}{E} = -\frac{dx}{X_0} \tag{2.1}$$

where X_0 is called the *radiation length* (of the medium) and can be calculated by using Equation 2.2:

$$\frac{1}{X_0} = \frac{4Z(Z+1)r_e^2 N_0}{137A} \ln\left(\frac{183}{Z^{1/3}}\right) \tag{2.2}$$

where $r_e^2 = e^2/mc^2$ is the classical electron radius. (For hydrogen $X_0 = 58\ g/cm^2$, while for lead, $X_0 = 6.5\ g/cm^2$). From Equation 2.1 it follows that the average energy of a beam of electrons of initial energy E_0, after traversing a thickness x of medium, will be:

$$< E >= E_0 \exp(-\frac{x}{X_0}) \tag{2.3}$$

The actual energy loss suffered by a given electron varies widely about the mean, and therefore it is not possible to speak of a well-defined range for a high-energy

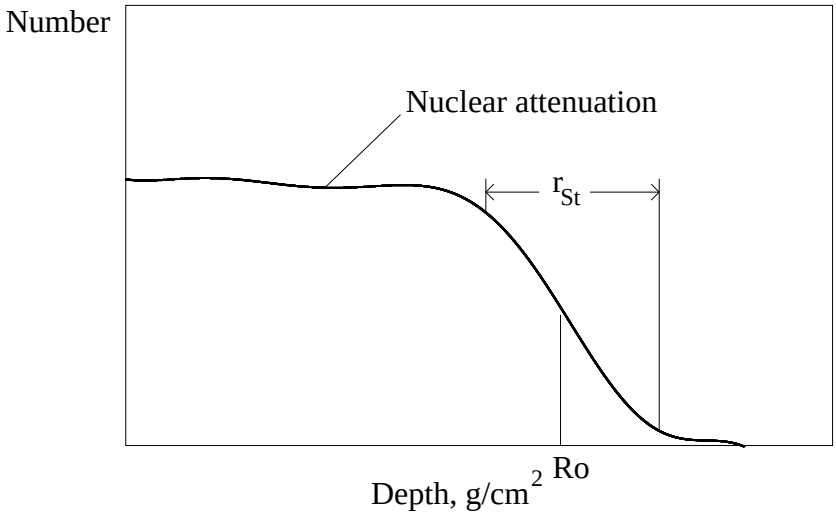

Figure 2.1: The variation of intensity of a beam of particles (initially monoenergetic) with depth. The particles gradually lose energy and eventually stop. (The depth in grams per square centimeter is the product of the distance traversed and the density of the material.)

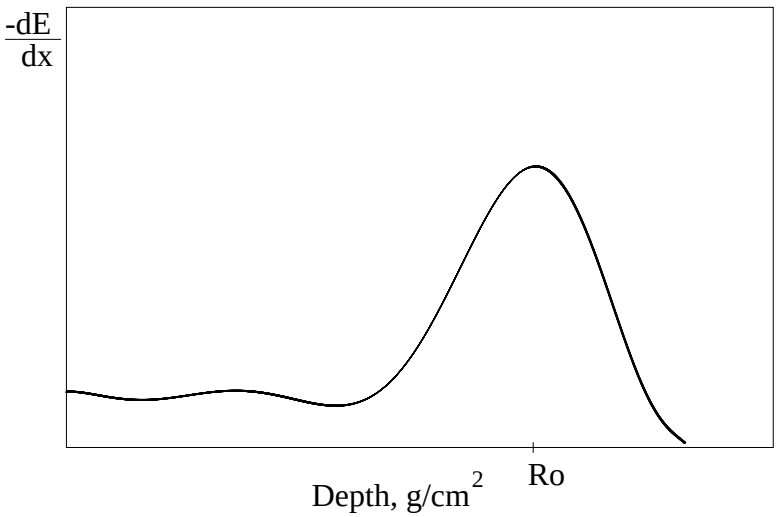

Figure 2.2: The variation in the rate of energy loss for a charged particle with depth. The rate of energy loss increases dramatically as the particle slows down near the end of its range.

electron in a high-Z material (for which radiation losses are much greater than collision losses).

2.3.3 Multiple Coulomb Scattering

When a charged particle passes through matter, it undergoes a series of scatterings because of encounters with electric fields of individual nuclei. This causes a deviation of the particle from its original direction. The deviation caused by a single encounter is likely to be small, but the cumulative effect of many encounters can be rather large. This has an important bearing on the precision attainable in measuring the angles or momenta of a particle, say a bubble chamber. An approximate but useful formula for the rms angle of deviation is:

$$< \theta^2 >^{1/2} \approx \frac{21MeV}{p\beta c}\sqrt{\frac{x}{X_0}} \tag{2.4}$$

where θ is the space angle between the original direction of the particle and its direction after traversing a thickness x of the material, X_0 is the radiation length, $\beta = v/c$ and p is the momentum of the particle in units of MeV/c.

2.3.4 Čerenkov Radiation

When a charged particle passes through a transparent medium with a velocity greater than that of light in the medium (that is, $\beta > 1/n$, where n is the index of refraction), it loses a small fraction of its energy in the form of *Čerenkov radiation*. This was discovered by the Russian physicist P.A. Čerenkov in 1934 and become the basis for a detector similar to the scintillation counter, whose advantage is that it responds only to particles moving faster than a certain velocity.

2.4 The interaction of High-Energy Photons with Matter

In passing through matter, high energy photons (gamma rays) can lose energy by several means:

(1) *Pair Production.* The formation of an electron-positron pair in the field of a nucleus.

(2) *Compton Scattering.* The elastic scattering of a photon by a free electron (or other charged particles).

(3) *Photo-electric Effect.* The process in which the photon gives up all its energy to a bound electron and removes it from the atom. Photo-electric absorption is typically strongly energy dependent, exhibiting a resonance-like behavior when the photon energy is just equal to that required to remove an electron from a given atomic shell.

For most elements photo-electric absorption is predominant for $E_\gamma \ll 1\ MeV$. Compton scattering is important for $E_\gamma \sim 1\ MeV$, and pair production predominates for $E_\gamma \gg 1\ MeV$. As a rule of thumb the mean free path for photon in material is X_0, radiation length previously defined. High-energy electrons or photons in matter typically initiate *showers* since the electrons produce photons by bremsstrahlung and photons in turn produce pairs. These processes are the basic operating principle of a *calorimeter* which will be explained later.

The process of conversion of a high-energy photon to an electron-positron pair (in the filed of a nucleus to conserve momentum) is closely related to that of electron bremsstrahlung. The attenuation of a beam of high-energy photons of intensity I_0 by pair production in a thickness x of absorber is described by:

$$I = I_0 \exp(-\frac{7x}{9X_0}) \tag{2.5}$$

so the intensity is reduced by a factor e in a distance $9X_0/7$, sometimes called the *conversion* length. It is also emphasized that, although the threshold for pair production is $E_{th} = 2m_e c^2 \approx 1\ MeV$, asymptotic value (2.5) for the absorption coefficient is not attained until one reaches photon energies of almost $1\ GeV$.

2.5 Particle Detectors

The detectors employed in experiments in high-energy physics are required to record the position, arrival time, and identity of charged particles. Precise evaluation of position coordinates is required to determine the particle trajectory and, in particular, its momentum (from the deflection in a magnetic field); precise timing is often required in order to associate one particle with another from the same interaction, frequently in situations where the total interaction rate per unit time may be very high. The identity of a particle may be established from simultaneous measurement of velocity (by time-of-flight or Čerenkov radiation) and momentum, and hence the rest mass; from the observation of decay modes, if the particle is unstable; and from its observed interaction with matter via strong, electromagnetic, or weak forces. Neutral particles are detected through their decay (e.g., $K^0 \rightarrow \pi^+\pi^-$) and/or interaction with matter (e.g., $\pi^0 \rightarrow \gamma + \gamma, \gamma \rightarrow e^- e^+$), leading to secondary charged particles.

There is no single detector to detect all kind of particles, so a combination of different detectors is employed.

2.5.1 Electroscope

Experimental study of elementary particles concerns, first of all, their production. In the early days of elementary particle research, radioactive sources, cosmic radiation and X-rays were important sources of high-energy particles. The main instrument for studying radioactivity and X-rays was the electroscope. This consists essentially of two thin gold leaves suspended side by side inside a glass and metal container. When the leaves are electrically charged a repulsive force is induced between them and causes them to separate. Now when the radioactive radiation or X-rays pass through the electroscope, the air molecules inside the container are ionised and air becomes a conductor. The charge on the leaves thus passes to the earthed container walls and drains away, causing the leaves collapse. The rate at which the leaves collapse is an indication of the intensity of the radiation.

2.5.2 Scintillation Counter and Photomultiplier

A more convenient and accurate instrument is *scintillator* or *scintillation* counter. One of its forms is a screen coated with fluorescent substance which emits light when an ionising strikes it. In its simplest form, a microscope is placed behind the scintillation screen and the observer counts the number of scintillations in a given time. In the more sophisticated version used today the task of detecting the light is performed by an electronic tube called a photomultiplier, which converts the scintillations into pulses of electricity; these are amplified and counted by electronic circuits. Modern scintillators are made of a transparent plastic block placed in contact with the photomultiplier. A particle passing through the plastic produces scintillations which reach the photomultiplier, are converted into the electronic pulses and recorded. Schematic of a photomultiplier tube can be seen in Figure 2.3.

2.5.3 Geiger-Müller Tube

A similar but more sensitive instrument, capable of recording a single particle, is the Geiger-Müller tube. The gas pressure and the voltage between the electrodes are such that as the ion produced by a single particle move towards the electrodes they are accelerated and ionize other atoms with which they collide, thus amplifying the current. So a single ionising particle produce a short,

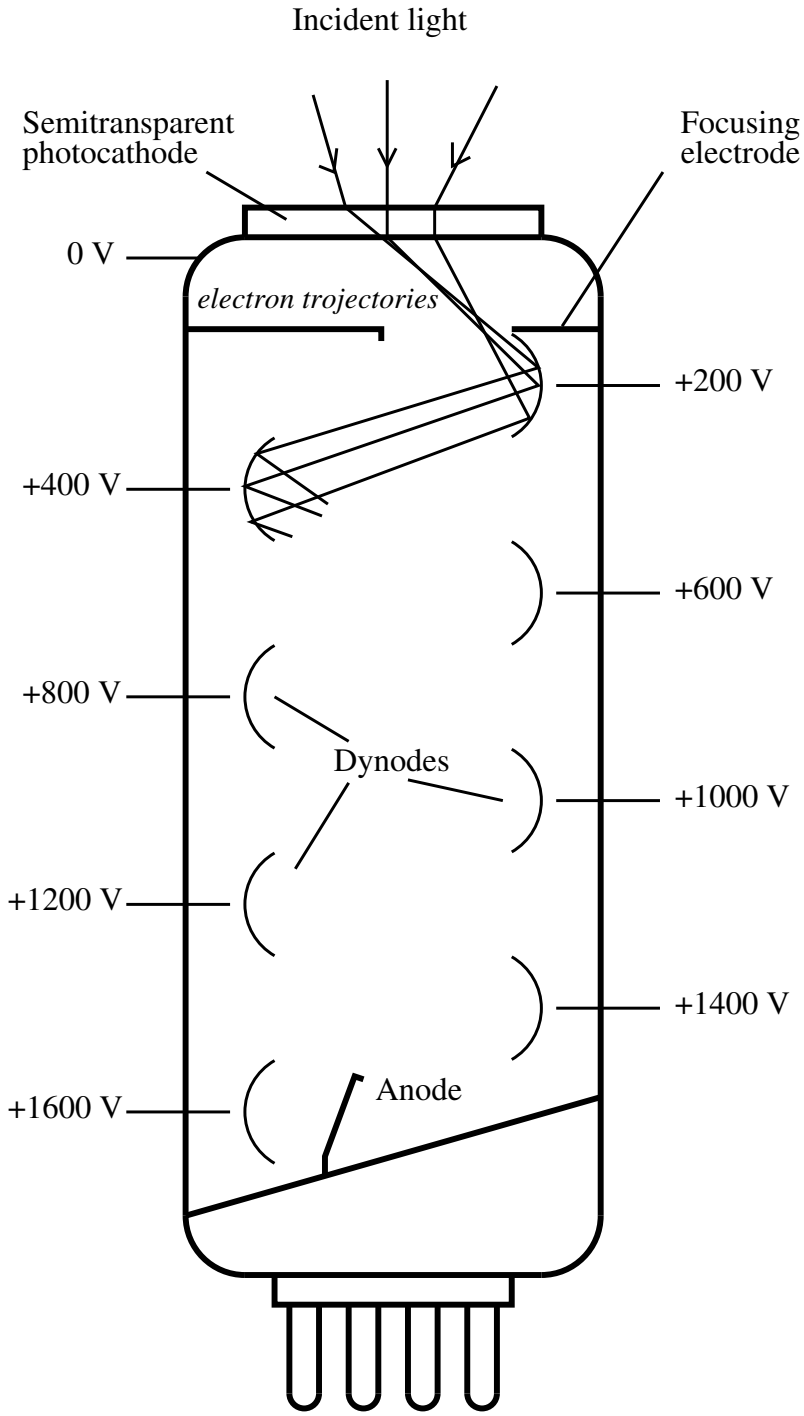

Figure 2.3: Schematic of a photomultiplier tube. Photomultipliers are invariably used to detect the light emitted from a scintillation or Čerenkov counter and convert it to an electrical pulse.

strong pulse. The number of pulses reflects the number of particles that have penetrated the tube. In the 1960s, solid state counters were developed in which the ionising particle produces a current in a semi-conductor crystal.

2.5.4 Proportional Counters

The proportional counter is one of the oldest devices for recording of ionisation. Single counters consist of a gas-filled cylindric metal or glass tube of radius r_2 maintained at negative potential, with a fine central anode wire of radius r_1 at positive potential. The electric field in the gas for a potential difference V_0 is then:

$$E(r) = \frac{V_0}{r \ln(r_2/r_1)} \tag{2.6}$$

An electron liberated by ionisation at radius r_a will drift toward the anode, gaining an energy $T = e\int_{r_b}^{r_a} E(r)dr$ when it reaches radius r_b. If T exceeds the ionisation energy of the gas, then fresh ions are liberated and a chain of such process leads to an avalanche of electrons and positive ions. The gas amplification factor, equal to the total number of secondary electrons reaching to the anode per initial ion pair, is typically $\sim 10^5$, independent of the number of primary ions; hence the name *proportional counter*.

The most significant advance in the field was the introduction of the multi-wire proportional counter (MWPC). This device consists of many parallel anode wires stretched in a plane between two cathode plates (Figure 2.4). This different anode wires act as independent detectors. A typical structure has wires of 20-μ m diameter with 2- mm spacing, between cathode planes 12 mm apart, operating at potential difference of 5 kV. In general, several electrons from the primary ionisation will drift toward an anode wire and create separate avalanches yielding negative pulses with very fast rise times (~ 0.1 ns). The positive ions have much lower mobility and induce pulses on both the cathode and neighboring anode wires of duration ~ 30 ns. The effective spatial resolution from the anode pulses is of order 0.7 mm.

2.5.5 Drift Chambers

As indicated above, the MWPC typically has spatial resolution of 1 mm or less, and time resolution of 30 ns. However, to achieve this resolution over large areas, an enormous number of wires (together with amplifiers) are required. A great reduction in cost can be achieved by drifting the electrons from the primary ionisation over 10 cm in a low-field region (1 kV/cm) before reaching the high-field amplification region near the the anode wire; the collection time of

avalanche then gives a measure of the position of the ionisation column. Figure 2.4 shows a typical arrangement of anode, cathode an field wires required to give a uniform drift field. Very large volume (100 m^3) drift chambers have been built. Sometimes such a device is called a *time projection chamber* (TPC).

2.5.6 Ionisation Chambers

In another type of instrument the ionising particle produces an electric current indirectly. One of the simplest instrument of this kind is the ionisation chamber. This is a closed, gas-filled chamber with two electrodes to which electric voltage is applied. A particle penetrating the chamber produces ions in the gas and these are drawn to electrodes (positive ions to the negative electrode and electrons to the positive electrode). An external circuit connected to the electrodes records the current, which is proportional to the radiation flux.

2.5.7 Bubble Chambers

The bubble chamber has been an indispensable tool of high-energy physics for many years, particularly studying complex interactions involving many secondary particles. Conceived by Glaser in 1952, it relies for its operation on fact that in a super-heated liquid[1] , boiling will start with formation of gas bubbles at nucleation centers in the liquid, and particularly along trails of ions left by the passage of a charged particle. The liquid filling is maintained under an overpressure (typically 5-20 atm), and super-heating achieved by sudden expansion of a piston, bellows, or diaphragm placed at the rear of the chamber. After expansion, bubbles along the tracks are allowed to grow over a period of order 10 ms and photographed by an array of stereo cameras using flash illumination. The bubbles then collapse under a recompression stroke. Since the cycle time is of order 1 s, the bubble chamber is well matched to pulsed, cyclic accelerators, with repetition rates of same order.

2.5.8 Electromagnetic Shower Detectors and Calorimeters

The energy and position coordinates of secondaries from high-energy interactions can also, under suitable conditions, be measured by total absorption methods. In the absorption process, the incident particle interacts in a large detector mass, generating secondary particles which in turn generate tertiary

[1] The most usual liquids filling for bubble chamber are hydrogen, deuterium and heavy liquids such as neon-hydrogen mixture, propane and Freon

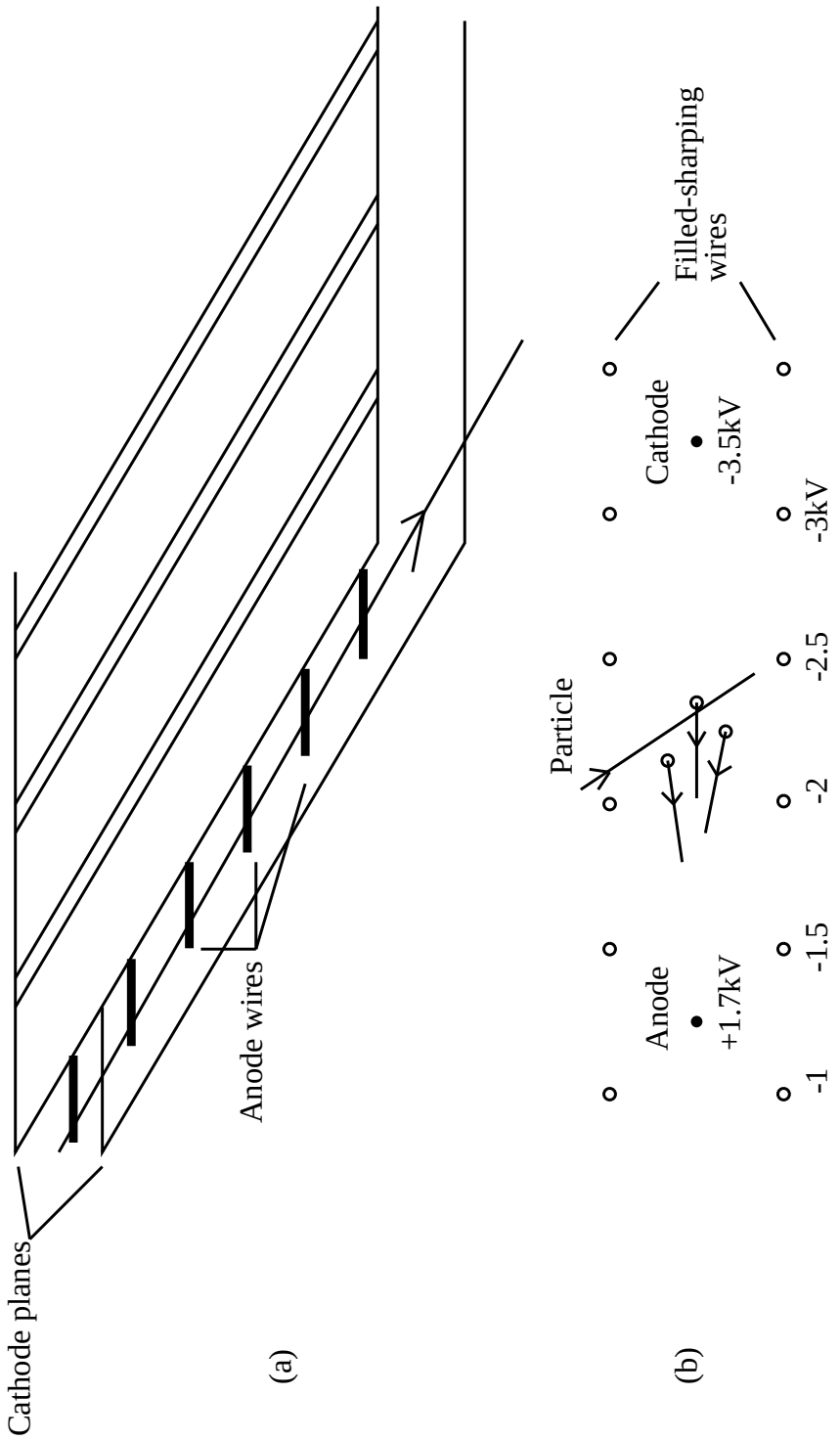

Figure 2.4: (a) Schematic layout of multiwire proportional chamber (b) Typical arrangement electrodes for a drift-chamber cell.

particles, and so on, so that all (or most) of the incident energy appears as ionisation or excitation in the medium-hence the term *calorimeter*.

For electrons and photons of high energy, a dramatic result of the combined phenomena of bremsstrahlung and pair production is occurance of cascade showers. A parent will radiate photons, which convert to pair, which radiate and produce fresh pairs in turn, the number of particle increasing exponentially with depth in the medium. A measurement of the position and charge of a shower provides a measure of the position an energy of the electron or photon.

Chapter 3

The ALEPH Detector

ALEPH (Apparatus for LEP pHysics) [2], was one of four detectors positioned around the LEP collider at CERN. It was in operation for six years in the study of e^+e^- collisions at Z resonance. The ALEPH project was the outcome of a collaboration of several hundred physicists and engineers from thirty universities and national laboratories, principally from Europe but also from China and the U.S.A.

ALEPH is a particle detector, covering as much of the 4π solid angle as possible (see Figure 3.1). It is designed to measure the momenta of charged particles, to measure the energy deposited in calorimeters by charged and neutral particles, to identify the three lepton flavors, and to measure the distance of travel of short-lived particles such as the tau lepton and the b and c hadrons. Particular emphasis has been given to momentum resolution up to the highest energies (by means of a large tracking system in a 1.7 T magnetic field), to electron identification (by means of a high segmented, projective electromagnetic calorimeter, as well as ionisation measurement in the tracing system), and muon identification (with continuous tracing inside sufficient iron absorber to eliminate the hadrons).

3.1 Overview

Figure 3.1 shows a schematic view of the ALEPH detector with the main features labeled. ALEPH uses a coordinate system usually expressed in terms of Cartesian (x, y, z) or cylindrical (r, ϕ, z) coordinates. In both cases, the z direction is along the beam line with the e^- direction defining the positive axis. The positive x direction points to the centre of LEP and the positive y direction is defined such that (x, y, z) forms a right-handed coordinate system.

The tracking system involves three sub-detectors: a silicon vertex detector, a drift chamber with 30 cm outer radius, which is also important as part of

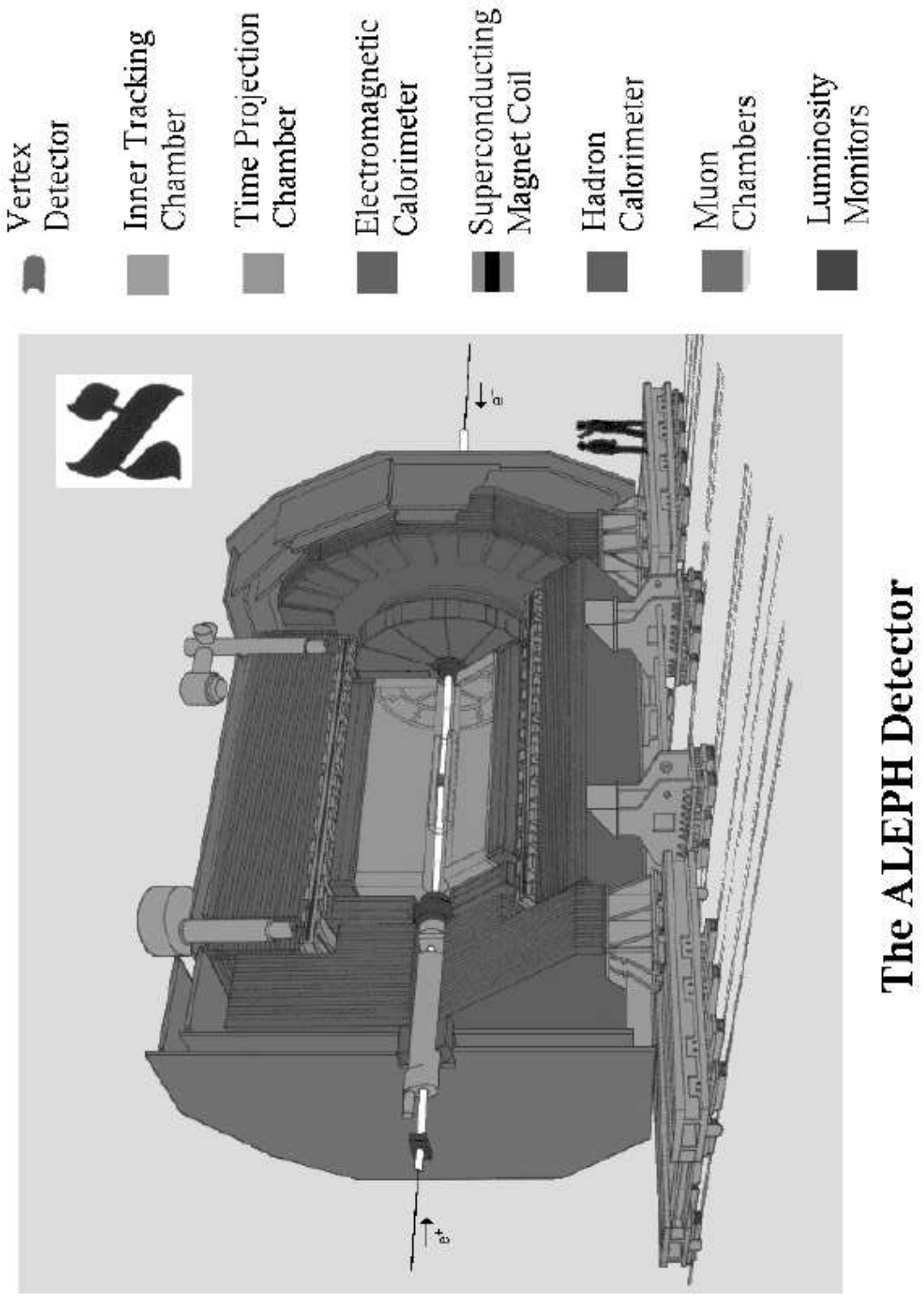

Figure 3.1: Schematic view of the ALEPH detector

the trigger system, and a time projection chamber with 180 cm outer radius. Calorimetry proceeds in two stages: electromagnetic and hadronic. A muon detection system of two double-layers of streamer tubes surrounds the whole detector. Finally, important for precise cross-section measurement are the highly segmented luminosity calorimeters.

Summary of the main features of the apparatus is presented here.

3.2 The Central Tracking Chambers

3.3 The Vertex Detector (VDET)

Close to the interaction point, tracking is performed by a silicon vertex detector. Vertex detector hits are used to provide additional precision for tracks already reconstructed in the outer tracking.

The device is formed from 96 silicon wafers each of dimensions ($5.12 \times 5.12 \times 0.03$) cm, arranged in two coaxial cylinders around the beam pipe. The spatial resolution, for normally incident tracks is: $\sigma_{r\phi} \approx 12\ \mu m$, $\sigma_z \approx 10\ \mu m$.

3.4 The Inner Tracking Chamber (ITC)

The vertex detector is surrounded by a conventional cylindrical multi-wire drift chamber. It measures the $r\phi$ position of a track on eight concentric layers of hexagonal drift cells at radii between 16 and 26 cm, with adjacent layers staggered by half a cell width. The resolution depends on the drift length in the cell, with an average of 150 μm. The position of tracks along the beam direction (z) is determined by measuring the difference in arrival time of the signals at each end of the wires. This, however, only has a resolution of a few centimetres. The ITC tracking information is used to improve tracking resolution, and to provide information for the first level trigger decision (made within 2-3 μs after a beam crossing).

3.5 The Time Projection Chamber (TPC)

The TPC is ALEPH's principal tracking detector providing 21 3D space-point measurements for fully contained tracks at radii between 30 and 180 cm. It consists of a gas volume (at atmospheric pressure) bounded by inner and outer field cages and two end-plates. An electric field is created between a central membrane and the two end-plates (dividing the TPC chamber into two), giving

a drift length of 2.2 m on each side. As charged particles pass through the chamber they leave a trail of ionized gas. The electrons drift along the electric field lines towards one of the end-plates where they induce ionisation avalanches in proportional wire chambers. These are detected as pulses on cathode pads. The ϕ coordinate is calculated by interpolating the signals induced on cathode pads and the r coordinate is given by the radial position of the pads involved in the measurement. The z coordinate of a point on a track's trajectory is obtained from the electron drift time and the drift velocity.

The five helix parameters shown in Figure 4.5 are determined by fitting a helix to the pad coordinates within the first half turn of each track candidate. A circle fit made in the xy plane yields ω (the inverse radius of curvature), d_0 (the impact parameter in the xy plane), and ϕ_0 (the emission angle in the xy plane). A straight line fit in the s_{xy}-z plane yields z_0 (the z coordinate at the closest approach to the z axis) and tanλ (the tangent of the dip angle). The circle fit parameters are used to calculate the values of s_{xy}.

The pulse height information on the wires provides information about the energy lost by the particle as it traverses the TPC. The rate of energy loss (dE/dx) is dependent on the particle mass and so can be used to identify different particle species. Electron identification is good, with greater than 3σ separation up to $p \approx 8\ GeV/c$. The π-K separation is roughly constant above $p \approx 2\ GeV/c$ at about 2σ, while the K-p separation is only about 1σ. Therefore, kaon and proton identification can be accomplished only on a statistical basis; nonetheless, it is an important means of reducing combinatorial background in many analyses. The azimuthal coordinated resolution is $\sigma_{r\phi} = 180\ \mu m$ at 0° pad crossing angle. The z spatial resolution for wires is $\sigma_z = 1.2\ mm$ (with a small z dependence), and 0.8 mm for pads (at $\theta = 90^\circ$).

The transverse momentum resolution and impact parameter resolution are shown in Table 3.1 for the TPC only, for the TPC and drift chamber, and for all three tracking detectors together. At low momentum (less than 0.4 GeV) a constant term of 0.5% should be added to the resolution due to multiple scattering.

3.6 The Electromagnetic Calorimeter (ECAL)

The ECAL is a sampling proportional wire calorimeter consisting of lead sheets and proportional wire chambers covering the angular range $|\cos\theta| < 0.98$. The calorimeter stops and measures the energy of electrons and photons. The

ECAL is formed from a barrel surrounding the TPC, closed at each end by an end-cap. These are divided into 12 modules, each covering an azimuthal angle of 30°. The cracks between the modules, where the ECAL is not sensitive to particles, constitute 2% of the barrel surface and 6% of the endcap surface. To ensure that the cracks in the endcaps and the barrel are not coincident the endcap modules are rotated through 15° azimuth. The modules have 45 lead/wire-chamber layers, with a total thickness of 22 radiation lengths.

The energy and position of each shower is read out using small cathode pads with dimensions $\sim (30 \times 30)\ mm$ arranged to form towers pointing to the interaction point; each tower is read out in three segments in depth, known as storeys, with thicknesses of 4, 9, and 9 radiation lengths respectively. There are 74,000 such towers, corresponding to an average granularity of $0.9° \times 0.9°$. This fine segmentation is important in the identification of photons, electrons and neutral pions. The angular resolution of the ECAL is:

$$\sigma_{\theta,\phi} = \left(\frac{2.5}{\sqrt{E}} + 0.25\right) mrad, \tag{3.1}$$

where E is measured in GeV.

The energy resolution, determined by comparing the measured energy to the track momentum or beam energy, is shown in Figure 3.2a as a function of the electron energy. The corresponding fitted resolution is:

$$\frac{\sigma_E}{E} = \frac{0.18}{\sqrt{E(GeV)}} + 0.009 \tag{3.2}$$

where a linear sum is used to as it gives the best fit; using a quadratic form would give:

$$\frac{\sigma_E}{E} = \frac{0.178}{\sqrt{E(GeV)}} \oplus 0.019 \tag{3.3}$$

The energy resolution as a function of polar angle is plotted in Figure 3.2b for electrons from Bhabha events. In the region around 40° in polar angle the

Table 3.1: Momentum and impact parameter resolution for the TPC, the TPC+ITC, and the TPC+ITC+VDET. At low momentum a constant term of 0.5% should be added to the momentum resolution due to multiple scattering.

	Transverse momentum resolution (p_T in GeV)	Impact parameter resolution	
Detector	$\Delta p_T/p_T$	$\sigma_{r\phi}(\mu m)$	$\sigma_{rz}(\mu m)$
TPC	$1.2 \times 10^{-3}\ p_T$	310	808
ITC	$0.8 \times 10^{-3}\ p_T$	107	808
VDET	$0.6 \times 10^{-3}\ p_T$	23	28

energy resolution is degraded, this corresponds to the overlap region where the electromagnetic shower develops into both barrel and end-cap modules. In that region the total thickness decreases, reaching only 16 radiation lengths, and in addition there is more uninstrumented material due to end-plates of modules and cables.

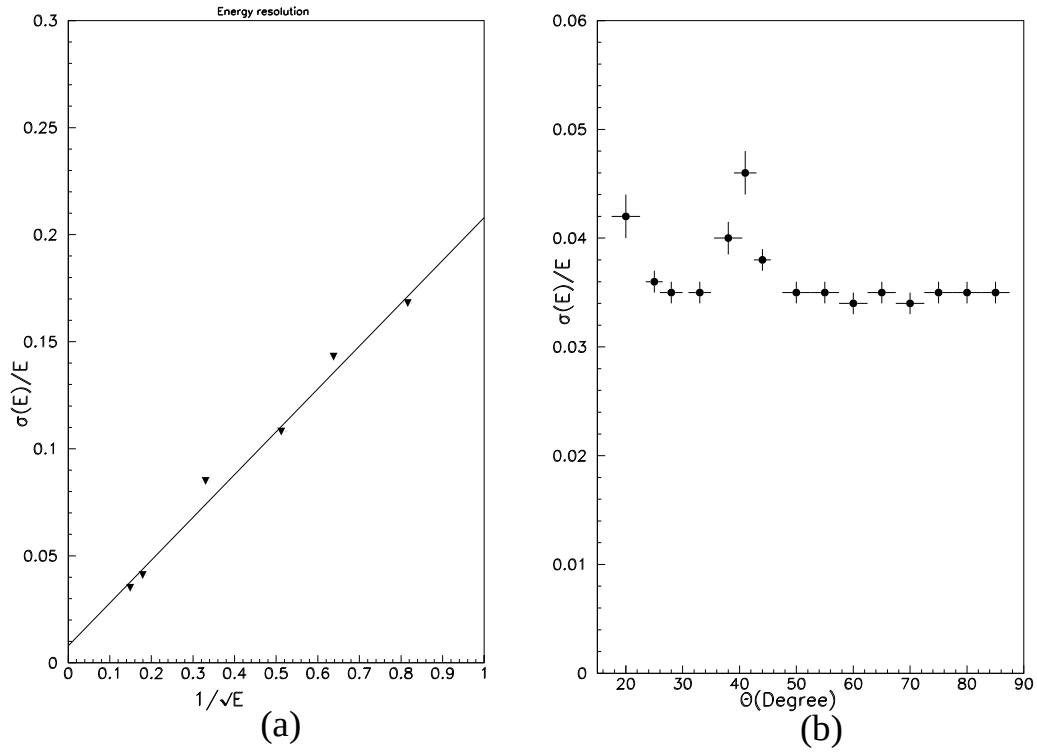

Figure 3.2: Energy resolution of the electromagnetic calorimeter; (a) dependence on energy, (b) dependence on polar angle for electrons from $Z \rightarrow e^+e^-$ decays.

3.7 The Hadron Calorimeter (HCAL) and Muon Chambers

The hadron calorimeter serves two purposes. It is used, together with the electromagnetic calorimeter, to measure hadronic energy deposits, and it is part of the muon identification system. It consists of 23 layers of plastic streamer tubes separated by 5 *cm* thick iron slabs, giving a total of 7.2 interaction lengths at 90 degrees. The calorimeter is constructed from 36 modules, 24 in the barrel and 6 in each end-cap, and is read out capacitatively in 4788 projective towers with a typical tower coverage of $3.7° \times 3.7°$, corresponding to 4×4 of the electromagnetic calorimeter towers. The streamer wire signals are summed and used as part of the trigger. The energy resolution of the hadron calorimeter for pions at normal

incidence is:

$$\frac{\Delta E}{E} = \frac{0.85}{\sqrt{E}} \tag{3.4}$$

where E is measured in GeV.

Outside the HCAL are another two double layers of streamer tubes which form the muon chambers (acting as simple tracking devices). Two coordinates are obtained from cathode strips both parallel and perpendicular to the wires in each layer. In conjunction with HCAL, the muon chambers are used to identify muons and keep hadron/muon misidentification to a minimum.

3.8 The Superconducting Solenoid

The ALEPH magnet is a liquid helium-cooled superconducting solenoid with a magnetic field strength of 1.5 T at a current of 5000 A. The field provides the curvature to charged tracks from which the momentum and sign of charged particles can be determined.

3.9 The Luminosity Monitors

The precise measurement of electroweak parameters requires accurate knowledge of the beam luminosity. The reaction rate for a process is given by R = $\sigma\mathcal{L}$ where σ is the interaction cross-section and $\mathcal{L}$ is the luminosity. Low angle elastic (Bhabha) scattering is used to measure the luminosity as it is almost a pure QED process with very little interference from the weak sector, and has a well known cross-section.

3.10 The Silicon Calorimeter

The Silicon Luminosity Calorimeter (SiCAL) is ALEPH's main detector for providing luminosity measurements. It is positioned around the beam pipe at each end of the detector and covers angles between 24 to 58 $mrad$ from the beam axis.

The SiCAL uses 12 silicon/tungsten layers (23.3 radiation lengths) to sample showers produced by Bhabhas. Finely spaced pads determine the angle of an incident particle with a polar resolution of 0.15 $mrad$. The shower depth gives a measure of the energy deposited with a resolution of 3.4% at 45.5 GeV.

3.11 The Luminosity Calorimeter

The Luminosity Calorimeter (LCAL) is a lead/wire calorimeter, similar in design to the ECAL, comprising 38 layers (24.6 radiation lengths). Modules are positioned, on either side of ALEPH, just behind the SiCAL modules. The LCAL monitors polar angles from 45 to 190 $mrad$ with a resolution of 0.5 $mrad$ (at 45.5 GeV), and has an energy resolution of $\sim 0.15/\sqrt{E} + 0.01$, where E is measured in GeV.

3.12 BCAL

Online luminosity monitoring is provided by the Bhabha calorimeter (BCAL). This is a sampling calorimeter made of layers of tungsten and plastic scintillator, with a single plane of silicon strips to provide position measurements. The BCAL is positioned next to the beam pipe at 7.7 m from the interaction point, receiving a Bhabha hit rate of ≈ 7 Hz.

Chapter 4

Event Reconstruction and Simulation

4.1 Introduction

Recorded data containing the raw event information are reconstructed offline by the JULIA program. Tracking and calorimeter reconstruction is performed on both real and simulated data (simulated data will be discussed in Section 4.5). The results from JULIA are formed into BOS banks and written to Production Output Tapes (POTs) and then to Data Summary Tapes (DSTs) ready for physics analysis; DSTs contain almost all the information from the POTs, but with non-interesting events (i.e. noise and background) removed. Figure 4.1 illustrates the flow of data from event production to the storage of reconstructed events ready for physics analysis.

A more compact form of data storage is the Mini-DST, which is important for analyses which process a large number of events. Mini-DSTs contain less information and the data are integerised, so that these tapes can be staged to disk and read by the analysis program more quickly than DSTs. Additionally, results of some time-consuming algorithms are already available on the Mini-DSTs, further reducing processing time. The Mini-DST is used in the analysis presented in this thesis.

The reconstruction of data by JULIA is described here with emphasis on track reconstruction in the TPC and γ/π^0reconstruction in the ECAL. The final section describes event simulation.

4.2 Track Reconstruction

Figure 4.2 shows examples of the reconstruction of tracks from coordinates in the TPC. Each track is the result of fitting a helix to chains of 3-dimensional space coordinates which are consistent with the trajectory of a charged particle. Note that the tracks are curved in the xy projection, but straight in the z pro-

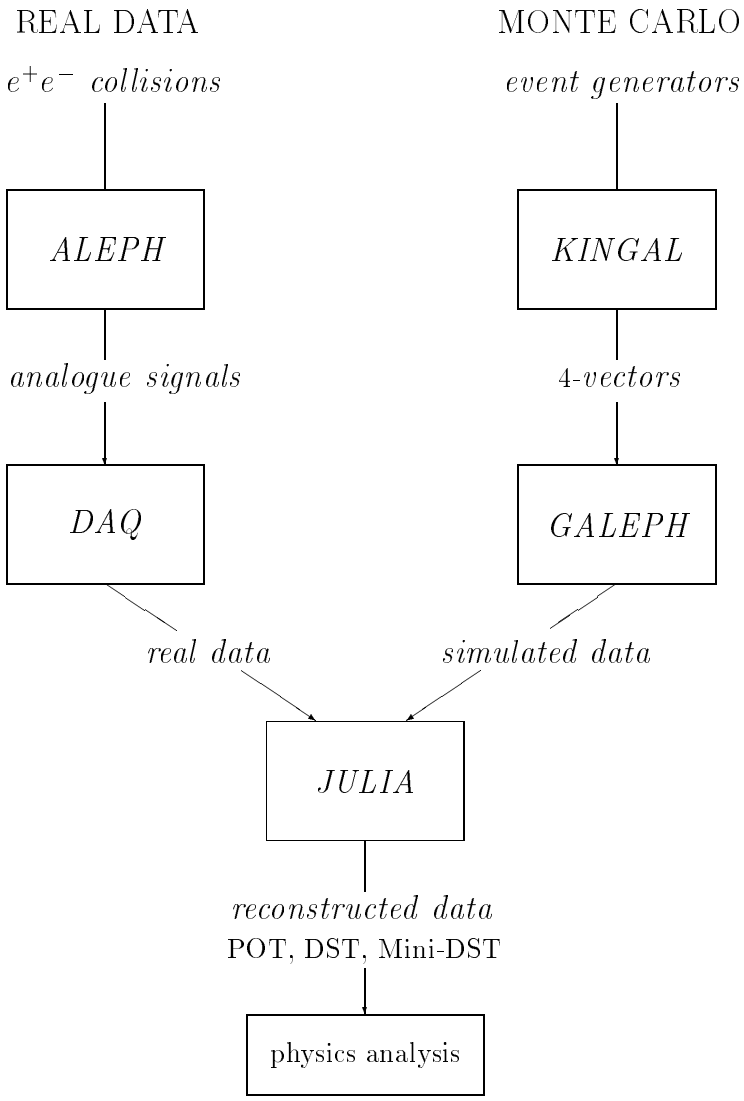

Figure 4.1: *The flow of data in the ALEPH experiment, including the simulated-data chain.*

jection. The determination of the coordinates, and the helix fitting is as follows.

4.2.1 TPC Coordinate Determination

The raw data collected from the TPC comes in two forms:

- The pad hits, which contain the pad number, arrival time and duration of the pulse.
- Digitized pulse-heights per time-slice, or 'buckets', corresponding to each pad hit.

Clusters and Pulses

Coordinate determination begins with the identification of clusters of pad hits (pulses) on a given pad row in the plane of 'pad number' versus 'drift time'. A cluster is built, ignoring for the moment pulse-height information, by starting with one hit and including hits on adjacent pads if they overlap the first by at least one sample. An example of such a cluster is shown in Figure 4.3, the cluster is formed by two nearby tracks which are not clearly distinguishable at this stage. A good cluster satisfies the following criteria:

- A minimum of 2, and a maximum of 20 pads in the cluster.
- The minimum cluster length is 5 and the maximum is 35.

A rejected cluster is flagged as a 'bad' coordinate; such clusters are not included in track reconstruction.

Subclusters and Subpulses

Once clusters have been established they are analysed again with the pulse-height information included in order to separate, or at least recognise, within each cluster the contributions coming from different particles. Figure 4.4 illustrates an example of a cluster which is separated, in drift time (or z), into three subclusters by inspecting the pulse-heights.

The charge profile of each pulse is analysed for evidence of multiple peaks. Peaks that are sufficiently isolated from others form subpulses. Groups of subpulses on adjacent pads are aligned with respect to time (or z) and grouped into subclusters. Within a subcluster, each subpulse must be on a separate pad. The charge profiles of the resulting subclusters are also scanned for evidence of multiple peaks; if a valley between peaks is sufficiently low then such a subcluster

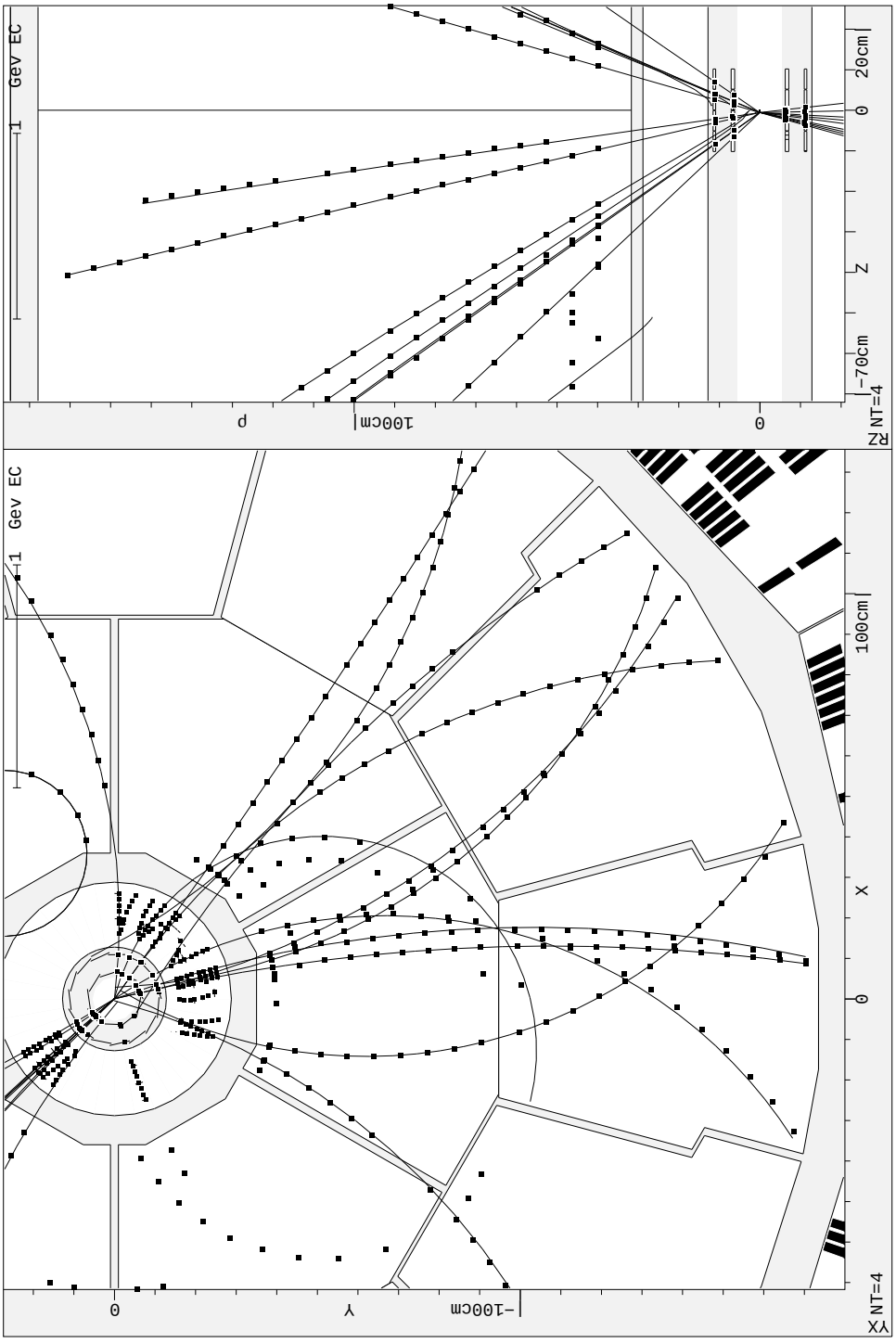

Figure 4.2: *Reconstruction of tracks from coordinates in the TPC. Two projections are shown: the xy projection, and the ρz ($\rho^2 = x^2 + y^2$) projection.*

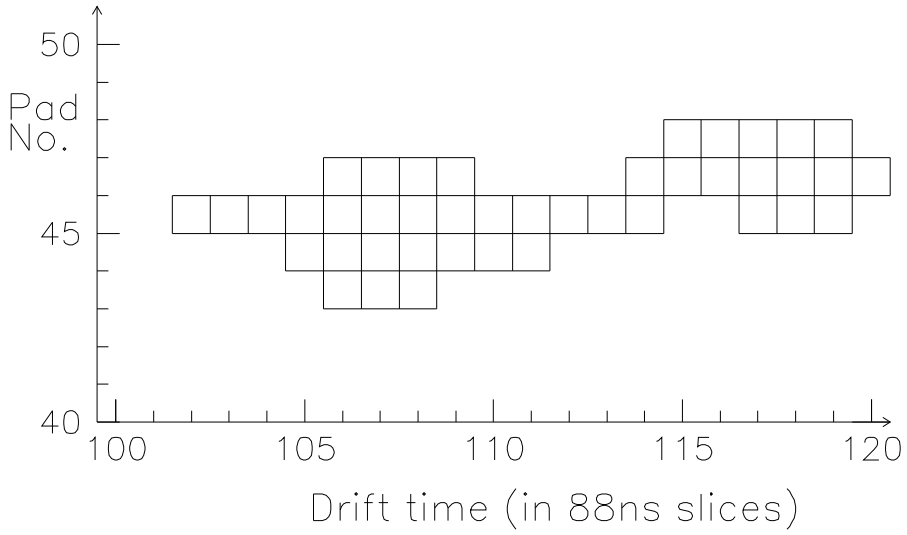

Figure 4.3: *A TPC pad row cluster formed by two nearby tracks.*

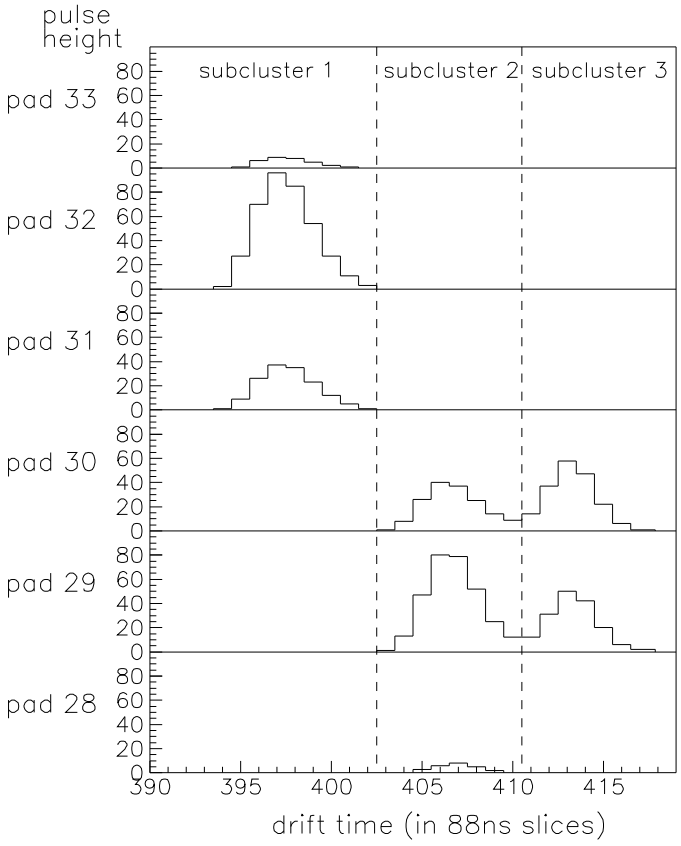

Figure 4.4: *An example of a complex cluster and how the reconstruction program has broken it down into three contributions from which three individual coordinates can be calculated.*

can be broken in $r\phi$, otherwise it is flagged as being unusable. Examples of such unusable clusters are shown circled in Figure 4.2: the hits from the two tracks originating from a photon conversion in the TPC gas are only resolved when the tracks are separated by a sufficient distance. For each subpulse, both a charge estimate and a time estimate are made from the digitisations. These are used in the reconstruction of coordinates.

4.2.2 Coordinates

Each 'good' subcluster is used to calculate a z and $r\phi$ coordinate. For the $r\phi$ coordinate the sub-pulse charge estimator is used: a Gaussian fit is made if the subcluster has 3 or less pads, or a charge-weighted average of the pad position is taken if the subcluster has more than 3 pads. The z coordinate is calculated using both the charge estimator and the time estimator.

4.2.3 TPC Track Finding and Fitting

Track candidates are formed from TPC coordinates as follows. Sets of coordinates which are consistent with an arc of a helix less than π radians are formed into 'chains'. Chains that are determined to belong to the same helix (e.g. multiple arcs of a spiral) are linked together into a single 'track candidate'). To account for multiple scattering, the coordinate error estimates are increased in accordance with the distance from the track origin. If an acceptable fit is not obtained, then up to two points are removed from the fit, corresponding to those that contribute the most to the χ^2 of the overall fit. If this does not result in an acceptable fit then a search is made for a kink in the track candidate, and if one is found the track is split. If no significant kink is found then a search is made for bad points by fitting a track candidate with one point at a time removed. If none of these methods leads to a good fit the track candidate is kept without modification. Details of the algorithms used to find and combine the chains into track candidates can be found in [3].

The track finding efficiency in the TPC has been studied using Monte Carlo simulation. In hadronic Z events, 98.6 % of tracks that cross at least four pad rows in the TPC are reconstructed successfully; the small inefficiency, due to track overlaps and cracks, is reproduced to better than 10^{-3} by the simulation.

The five helix parameters shown in Figure 4.5 are determined by fitting a helix to the pad coordinates within the first half turn of each track candidate.

A circle fit made in the xy plane yields ω (the inverse radius of curvature), d_0 (the impact parameter in the xy plane), and ϕ_0 (the emission angle in the xy plane). A straight line fit in the s_{xy}-z plane yields z_0 (the z coordinate at the

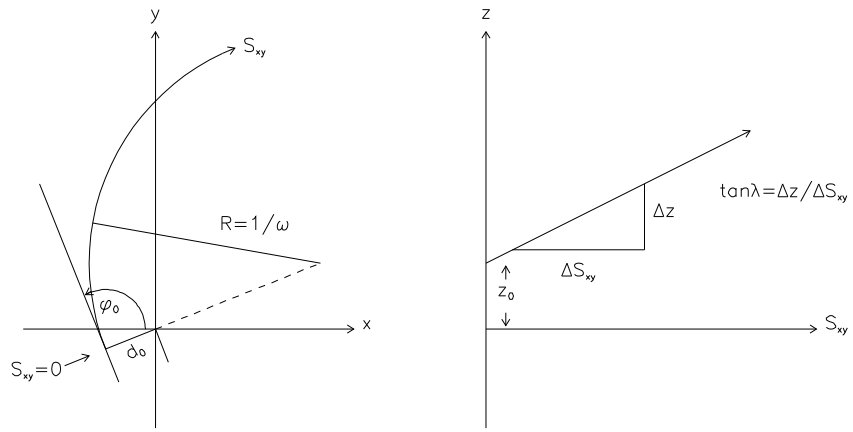

Figure 4.5: *Helix parameters used in the TPC tracking algorithms. For this case all parameters are positive.*

closest approach to the z axis) and tanλ (the tangent of the dip angle). The circle fit parameters are used to calculate the values of s_{xy}.

4.2.4 TPC-ITC-VDET Track Association

Once TPC tracks have been identified, ITC and VDET coordinates are added to improve the track measurement. The TPC tracks are projected back into the ITC and a search is made for ITC coordinates within a 'window' defined in z and ϕ around the trajectory. If no ITC hits are found in the first two outer layers then the search is abandoned. If more than three hits are found, a fit is performed and the ITC coordinates may be accepted, subject to a cut on the χ^2 of the fit. Finally VDET hits are associated to tracks which extrapolate close to them and the tracks are then refitted.

4.2.5 Ionisation Energy Loss

The TPC also provides information on a charged particle's rate of loss of energy due to ionisation (dE/dx). This information is obtained from the measured ionisation on the TPC sense wires and is calculated as follows:

(1) The raw digitisations for each wire pulse are reduced to time and charge estimates.

(2) The wire pulses are associated with the reconstructed TPC tracks. Individual wire pulses that match to more than one track are ignored, as are pulses which do not have a shape consistent with single track ionisation.

(3) The truncated mean dE/dx is calculated for each track. The dE/dx samples are distributed according to a Landau distribution and the mean

is calculated after discarding the upper 40% and lower 8% of the sample. The upper truncation is to reject samples arising from hard scattering, and the lower one reduces the dependence of the dE/dx estimator on track angle and drift length.

4.3 Calorimeter Reconstruction

Data from the ECAL and HCAL are also processed by JULIA. This section describes briefly the formation of calorimeter objects, the following section describes in more detail the reconstruction and identification of photons in the ECAL.

4.3.1 Calorimeter Clusters and Objects

Clusters of energy deposition in the ECAL are formed by associating a group of spatially connected storeys (i.e. storeys which possess a common corner). All storeys which have an energy exceeding the threshold t_{low} (typically set to 30 MeV to minimize effects of noise) are used as input to the cluster finding algorithms. A topological cluster is only retained if it contains at least one storey above the threshold t_{high} (typically 90 MeV).

ECAL clusters are then associated with nearby charged tracks, clusters which are more than 3 cm (about one pad width) from the entry point of a charged track into the ECAL are considered to be neutral. Clusters associated with minimum ionizing tracks, electrons, and neutral particles are distinguished. Charged clusters with excess energy may have photons associated with them. A similar process is carried out to form HCAL clusters. Calorimeter objects are then formed from the calorimeter clusters with the possibility of more than one particle contributing to each object. ECAL and HCAL objects which overlap are merged into combined calorimeter objects. Further information on calorimeter reconstruction can be found in [3].

4.4 γ/π^0Reconstruction and Identification

The clustering algorithm described above builds large clusters, often merging energy from photons and hadronic interactions. For the purpose of improving photon reconstruction, particularly in hadronic events, an extension to the above algorithm is implemented. It uses the facts that electromagnetic showers generally start in the first segment in depth of the electromagnetic calorimeter and that, unlike the cell patterns of hadronic clusters, storeys receiving energy from a

photon have a compact arrangement and most of them share a face with another storey associated to the same photon. Figure 4.6 shows the energy deposits of three photons in the ECAL, the showers begin in stack 1 (the first picture), they have a maximum in stack 2, and have almost completely disappeared by stack 3. It is apparent from the figure that the large cluster contains two subclusters due to two photons, the algorithm for photon reconstruction described here attempts to resolve these clusters by forming 'G-clusters' as follows.

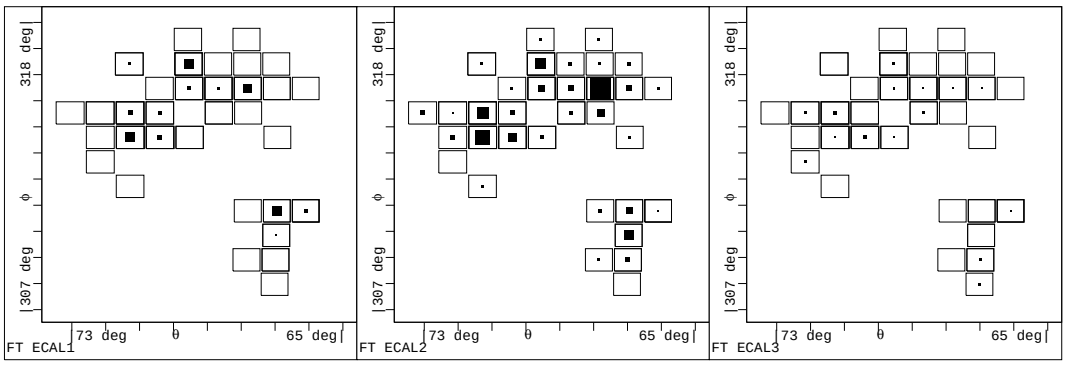

Figure 4.6: *Energy deposits of three photons in ECAL towers; the showers begin in stack 1 (the first picture), they have a maximum in stack 2 and almost completely disappear by stack 3. All towers contributing to the cluster are shown in each stack; empty boxes represent tower storeys containing zero energy.*

4.4.1 G-clusters

G-clusters are formed as follows. The storeys of stack 1 of the ECAL are scanned in the order of decreasing energy. A storey without a more energetic neighbour defines a new G-cluster. Other storeys in the same stack are assigned to the cluster of their highest energy neighbour, storeys are considered neighbours only when they share a common face. The procedure is repeated for storeys in the second and third stacks but, when processing a storey, the algorithm looks first for a neighbour in the previous stack. G-clusters found by the algorithm are retained as photon candidates only if their energy is greater than 0.25 GeV,and if there is no charged track impact at a distance of less than 2 cm from the G-cluster barycentre. The efficiency and background depend strongly on the density of particle impacts on the calorimeter.

4.4.2 Photon Position Calculation

The position of the photon impact point is given by the G-cluster barycentre i.e. the energy-weighted mean of the coordinates of each storey centre. This position is corrected for biases due to the finite size of the calorimeter cells.

4.4.3 Photon Energy Calculation

In order to reduce the sensitivity of the energy measurement to hadronic background and clustering effects, the photon energy is computed from the energy collected in the four central towers of the cluster, and the expected value of the fraction of energy in the four towers, F_4. This fraction is computed from the parameterisation of the shower shape for a single photon in the calorimeter. The computation takes into account the calorimeter pad area and the distance between the photon impact and the nearest tower corner, as well as the variation with energy of the expected F_4. Corrections to the energy are computed for energy losses before and after the ECAL and energy loss in the region where the ECAL barrel and endcaps overlap.

4.4.4 π^0Reconstruction

The reconstruction of neutral pions from G-clusters provides an improvement in purity and efficiency. The identification of subclusters is particularly useful for higher energy pions where the energy deposits of the daughter photons overlap. Discussion of π^0reconstruction from G-cluster photons will be reserved until Appendix A where the methods will be shown in detail as part of the analysis presented in this thesis.

4.4.5 Mass Constraint

To improve the momentum resolution of the pion the reconstructed mass can be constrained (see Appendix B) to the nominal pion mass of 135 MeV/c^2. This is done by modifying the photon energies until the required mass is obtained; see figure 4.7. The directions of the photons are not changed as they are well measured.

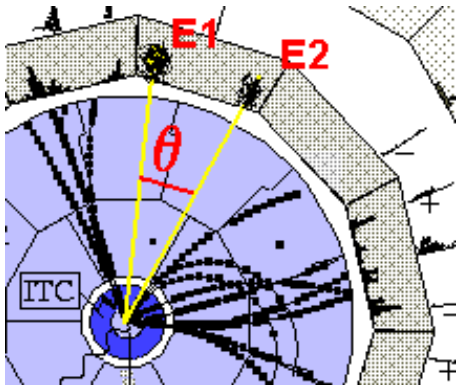

Figure 4.7: π^0*Decay into two photon*

4.4.6 Merged Photons

The probability of resolving a π^0into two G-clusters decreases for energies greater than $10GeV$. Further information is contained in the distribution, within a single cluster, of the energy in the calorimeter. To extract it, energy weighted moments of the two dimensional energy distribution are computed. Assuming that the cluster contains only two photons, by using moments up to the third order it is possible to reconstruct the two photon momenta and compute their effective mass. This allows a check of the π^0hypothesis for unresolved high energy clusters. An example of the use of this method can be found in [4], but it carries little advantage for the analysis presented in this thesis and so is not used here.

4.5 Event Simulation

Event simulation proceeds in two steps: simulation of the physics under investigation (carried out by event generators and the KINGAL program), and simulation of the response of the detector to the KINGAL particles (carried out by the GALEPH program).

4.5.1 KINGAL

The KINGAL package is effectively an interface between the event generators and GALEPH. The event generators use Monte Carlo simulation techniques to produce the desired physics. The KINGAL output gives a 4-vector kinematic representation of each simulated particle. Short-lived resonances are decayed according to decay tables based on experimental data; longer-lived particles are retained so that secondary vertices can be generated in GALEPH.

4.5.2 GALEPH

The Generator for ALEPH (GALEPH) is a detector simulation, representing the interaction of particles with the material of the detector, and the response of the detector to the particles. It generates digitized pulses with the same format as those produced by the data acquisition system.

The GALEPH package uses the GHEISHA package to simulate hadronic showers. Electromagnetic shower simulation is carried out using the GEANT software [5]. Particle scattering as well as energy loss calculations are also carried out by the GEANT package.

Chapter 5

Event Selection Using Adaptive Gaussian Kernels

5.1 Introduction

The Probability Density Estimation (PDE) method was implemented by researchers at Rice University for the top quark search [6]. In that analysis, the efficiency of the fixed kernel estimation was found to be similar to that of neural networks. Unlike neural networks, this method has few free parameters, allowing for less complicated optimization. Given a training sample of data, consisting of sets of signal and background, two functions of the N input variables are formed of the data set. This is accomplished by forming a product of kernel functions in the space of the input variables for each data point. The complete function for the entire sample of events is the sum of all of these product kernels in the training data, normalized by the number of training events used to form the function. This is known as the feature function, because it is an estimate of the important features in the data. Mathematically this function is given at Equation 5.4

5.2 Kernel Estimation

The notion of a kernel estimator grew out of the asymptotic limit of Averaged Shifted Histograms (ASH) [7]. The ASH is a simple device that reduces the binning effects of traditional histograms. The ASH algorithm is as follows: First, create a family of N histograms, $\{h_i\}$, with bin-width h, such that the first bin of the i^{th} histogram is placed at $x_0 + ih/N$. Because x_0 is an artificial parameter, each of the h_i is an equally good approximation of the parent distribution. Thus, an obvious estimate of the parent distribution is simply the average of the H_i, hence the name 'Average Shifted Histogram'. Note that resulting estimate

(with N times more bins than the original) is not a true histogram, because the height of a 'bin' is not necessarily equal to the number of events falling in that bin. However, it is a superior estimate of the parent distribution, because the dependence of initial bin position is essentially removed. In the limit $N \to \infty$ the ASH is equivalent to placing a triangular shaped kernel of probability about each data point t_i .

5.2.1 Fixed Kernel Estimation

In the univariate case, the general kernel estimate of the parent distribution is given by

$$f_0(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - t_i}{h}) \tag{5.1}$$

where $\{t_i\}$ represents the data and h is the smoothing parameter (also called the bandwidth). Immediately we can see that our estimate f_0 is bin-independent regardless of our choice of K. The role of K is to spread out the contribution of each data point in our estimate of the parent distribution. An obvious and natural choice of K is a Gaussian with $\mu = 0$ and $\sigma = 1$:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{5.2}$$

Though there are many choices of K, Gaussian kernels enjoy the attributes of being positive definite, infinitely differentiable, and defined on an infinite support. For physicists this means that our estimate f_0 is smooth and well-behaved in the tails.

Now we concern ourselves with the choice of the bandwidth h. In Equation 5.1 the bandwidth is constant for all i. Thus, f_0 is referred to as the fixed kernel estimate. The role of h is to set the scale for our kernels. Because the kernel method is a non-parametric method, h is completely specified by our data set $\{t_i\}$. In the limit of a large amount ($N \to \infty$) of normally distributed data , the mean integrated squared error of f_0 is minimized when

$$h^* = (\frac{4}{3})^{\frac{1}{5}} \sigma n^{\frac{-1}{5}} \tag{5.3}$$

Of course, we rarely deal with normally distributed data, and, unfortunately, the optimal bandwidth h is not known in general. In the case of highly bimodal data (e.g. the output of a neural network discriminate), the standard deviation of the data is not a good measure for the scale of the true structure of the distribution.

5.3 Multivariate Kernel Estimation

The general kernel estimation technique generalizes to d-dimensions . One choice for the d-dimensional kernel is simply a product of univariate kernels with independent smoothing parameters.

5.3.1 Fixed Kernel Estimation

For product kernels, the fixed kernel estimate is given by

$$\widehat{f_0}(\overrightarrow{x}) = \frac{1}{nh_1...h_d} \sum_{i=1}^{n} \left[\prod_{j=1}^{d} K(\frac{x_j - x_{ij}}{h_j}) \right] \tag{5.4}$$

The x are the set of variables this analysis is being performed on after transformation into a set in which correlations are zero (the new set of variables is a linear combination of the original variables). This transformation is done so that the kernel structure will match the covariance structure of the data, and thus give a better representation of the data points. Here, K is the kernel chosen to suit the data. In this analysis a Gaussian kernel has been chosen by means of Equation 5.2

The remaining items in the feature function are N_{tr} the number of training events, and $h_j = h^0 \times \sigma_j$, where σ_j is the standard deviation of variable j (in the space in which the correlations vanish), d is the number of variables used in the analysis and h^0 is a tunable parameter which must be optimized for the data set. h^0 is chosen such that the functions formed by the sum of the product kernels are smooth representations without losing information about the data. Using a linearly independent set of testing data, and the feature functions (f_s, f_b), a discriminant function value ($D(x)$) can be found for each event in the testing sample, representing a measure of how well the event matches the feature function for signal ($D(x)$ close to one) or background($D(x)$ close to zero). This discriminant function is simply:

$$D(x) = \frac{f_s}{f_s + f_b} \tag{5.5}$$

where f_s is the feature function for the signal and f_b is the signal function for the background. Then a single cut on the value of $D(x)$ can be made, selecting the signal from the background by the value calculated for each event.

5.3.2 Adaptive Probability Density Estimation

The PDE adaptive kernel builds on this method with one modification. An additional parameter α is used to further fit the Gaussian kernels to the data

set. A pilot $f(x)$ is found using h_j and then the analysis is performed using;

$$h_j = h^0 \times \sigma_j \times (\frac{f_{pilot(\bar{x})}}{f_{pilot(x)}})^{\alpha} \tag{5.6}$$

where α is another free parameter which must be optimized along with h^0 for the data set, f_{pilot} is the feature function formed using $h_j = h^0 \times \sigma_j$, and $\overline{x}$ is a vector of the mean values of each of the variables chosen for the analysis. This new choice of the width of the product kernels ties the functional form closer to the actual distribution of the data. For sets of data which may have a few outlying events, this now selects a wider Gaussian, so that the feature function is smoother for these areas.

Chapter 6

Comparison of PDE Methods

6.1 PDE Methods

The PDE attempts to describe the signal probability, or purity, of a distribution that contains both signal and background data. The signal probability $P(\vec{x})$ for a multi-variate data $\vec{x}$ is shown in Equation 6.1 where $f_S(\vec{x})$ and $f_B(\vec{x})$ are estimates of the underlying feature functions of the signal and background distributions respectively.

$$P(\vec{x}) = \frac{f_S(\vec{x})}{f_S(\vec{x}) + f_B(\vec{x})} \tag{6.1}$$

Here the data is obtained from Monte Carlo simulations of the processes that are under investigation.

The function $P(\vec{x})$ is used in the selection or rejection of events by applying a purity cut, P_{cut}: events with $P(\vec{x}) \geq P_{cut}$ (high signal probability) are accepted, and events with $P(\vec{x}) < P_{cut}$ are rejected. The value of P_{cut} is optimised with respect to a performance measure which in this study is chosen to be the product of the selected signal purity and efficiency, $\varepsilon \times \mathcal{P}$, and optimal is defined by the maximisation of this quantity. At least one other parameter, specific to the method used to form the feature function estimates, is involved in the optimisation of the PDE, optimisation therefore involves a parameter search in at least two-dimensions. To avoid biases in the training data (over-training), the function $P(\vec{x})$ is formed using a training data set, and the optimisation is performed using an independent test data set.

In high energy physics, Monte-Carlo simulations generally involve time-consuming modeling of signal and background processes, and detector simulations. Therefore, because of computing constraints, training and test data are limited in statistics. If statistics are not sufficient, the resultant functions $f_S(\vec{x})$ and $f_B(\vec{x})$ will be poor estimates leading to an inaccurate PDE and therefore less than optimal discrimination performance. The problem of limited statistics

becomes greater as the dimensionality of the data increases, the so-called 'curse of dimensionality'.

To combat the problem of statistics, a PDE method employs an algorithm for the formation of the feature function estimates such that the data is generalised, or in other words, smoothed. However, over-smoothing the data leads to loss of detail, the algorithm and its optimisation must be chosen carefully. A number of such algorithms are summarised in the following sub-sections.

6.1.1 The Histogrammed PDE, HPDE

The histogram is the traditional algorithm for representing the distribution of data. In a simple histogramming algorithm, the signal and background data are binned to create multi-dimensional histogrammed functions $f_S(\vec{x})$ and $f_B(\vec{x})$. Training of the PDE simply involves the formation of the histograms $f_S(\vec{x})$ and $f_B(\vec{x})$, and then a PDE histogram formed by dividing the histogrammed functions according to Equation 6.1. Optimisation of the PDE is performed with respect to the binning: the histogramming procedure is repeated for different total number of bins N and the binning that gives the best PDE performance is chosen. This procedure can be performed with a relatively small number of iterations by searching through values of N in the sequence: $N = 2^i$ with $i = 1, 2, 3, 4, ..., 16$ (a total number of bins $N = 2^{16} = 65536$ should be enough for most data, the value can be increased if required). The time taken to train and optimise the PDE is proportional to the number of data n, i.e. the time complexity of the algorithm is $O(n)$.

The optimal number of bins represents a balance between large N that allows the histogram to model the form of the data though with large statistical bin-to-bin fluctuations, and small N that reduces (smoothes) bin-to-bin fluctuations but at the possible expense of losing the underlying form of the data. At low statistics, or high dimensions, a good balance might not be achievable resulting in a poor PDE performance. The requirement that the data be bounded by upper and lower limits of the histogram can also give rise to performance problems when statistics are low.

6.1.2 The Smoothed Histogrammed PDE, SHPDE

The simple histogramming method can be improved by, for example, applying a smoothing procedure to the histograms, or applying variable bin width to increase binning in regions of high statistics and high gradients, and reduce binning in regions of low statistics. To demonstrate the improvements that can be gained, a simple Laplace histogram smoothing algorithm (averaging of neigh-

boring bins) is employed: a mask containing three consecutive bins (multiplied by each dimension) is scanned along the histogram, the value of the central bin of the mask is replaced with the mean of all values in the mask, in order to conserve the total number of entries in the histogram, the change in the value of the central bin is divided and subtracted equally from each member of the mask, all changes in the mask are reversed if any negative values result, this procedure is repeated five times.

The result of this procedure is the smoothing of bin-to-bin fluctuations; this enables a larger number of bins to be used in cases where statistics are limited.

6.1.3 The Kernel PDE, KPDE

Smoothing of the data in the kernel PDE method[8, 7] is obtained by replacing each data point with a multi-variate Gaussian kernel; the width of the kernel is optimised with respect to the PDE performance measure. The functions $f_S(\vec{x})$ and $f_B(\vec{x})$ are built by summing all Gaussian kernel functions in the training data. During optimisation the computation of functions $f_S(\vec{x})$ and $f_B(\vec{x})$ is repeated for every point in the test data; the time complexity of the KPDE algorithm is therefore $O(n^2)$. In its simplest form the KPDE is simple to implement, and has the advantage of not requiring a knowledge of the range of the data. Improvements to the KPDE performance can be gained using a variable (adaptive) kernel width[6] such that a wider kernel is used in regions of low statistics.

6.1.4 The Histogrammed Kernel PDE, HKPDE

In an attempt to reduce the run-time of the KPDE method, the functions $f_S(\vec{x})$ and $f_B(\vec{x})$ may be histogrammed prior to the optimisation phase thereby reducing the time complexity of the algorithm to $O(n)$. However, this technique is only useful at high statistics, and inherits some of the disadvantages of histogramming: the range of the data needs to be defined, and some uncertainty in the PDE values is introduced. Additional optimisation of the binning is also required.

6.1.5 The Range-Search PDE, RSPDE

Smoothing of the data in the Range-Search PDE method[9] is obtained by counting signal and background data in a volume around the point of interest. The volume shape may be a hyper-cube, or ellipsoid, with dimensions typically taken from the range, or the RMS of the data, in each dimension. The scale of the

volume is optimised with respect to PDE performance. Counting is performed in the training data while the 'points of interest' are taken from the test data. As for the KPDE algorithm, the time complexity scales as $O(n^2)$. In its simplest form the RSPDE is simple to implement, and again, has the advantage of not requiring a knowledge of the range of the data. In Reference [9] the run-time is reduced by storing events in two multi-dimensional binary trees, though this is achieved at the expense of a much more complicated algorithm.

The above five PDE methods are summarised together in the flow chart given in Figure 6.1

6.2 PDE Performances

The performance of the above five PDE methods, in their simple forms, are compared using data from a sample of 250,000 hadronic decays of the Z boson, generated with the JETSET 7.4[10] Monte-Carlo. The events are passed through the ALEPH detector simulation and reconstruction program. Reconstructed photons are selected and combined in pairs to create pion candidates for the decay $\pi^0 \to \gamma\gamma$. Details of this data in the context of π^0 selection can be found in the study of the reaction $\omega \to \pi^+\pi^-\pi^0$[11].

In this comparison, the discrimination of pion signal from its background is optimised such that the product of the selected signal efficiency and purity is maximised. The pion candidates have already been selected from an invariant mass window giving an initial $\varepsilon \times \mathcal{P}$ value of 0.32. To improve further the discrimination between signal and background, three-dimensional PDEs are formed from the following discriminators: the χ^2 from a constraint of the mass of a candidate to the nominal π^0 mass, the angle θ between the daughter photons in the lab frame, and the energy E of the pion candidate. The distributions of these discriminators are shown in Figure 6.2.

For this study the data is separated into forty data sets each containing 500,000 pion candidates. Training is performed on one data set, and, optimisation and selection is performed on a second independent data set. The procedure is repeated twenty times to obtain mean values and their standard deviations. To obtain results for both low and high statistics, the total number of signal candidates n is varied from 50 to 500,000 according to $n = 5 \times 10^d$ where d is a data scale that takes the values 1, 2, 3, 4 and 5.

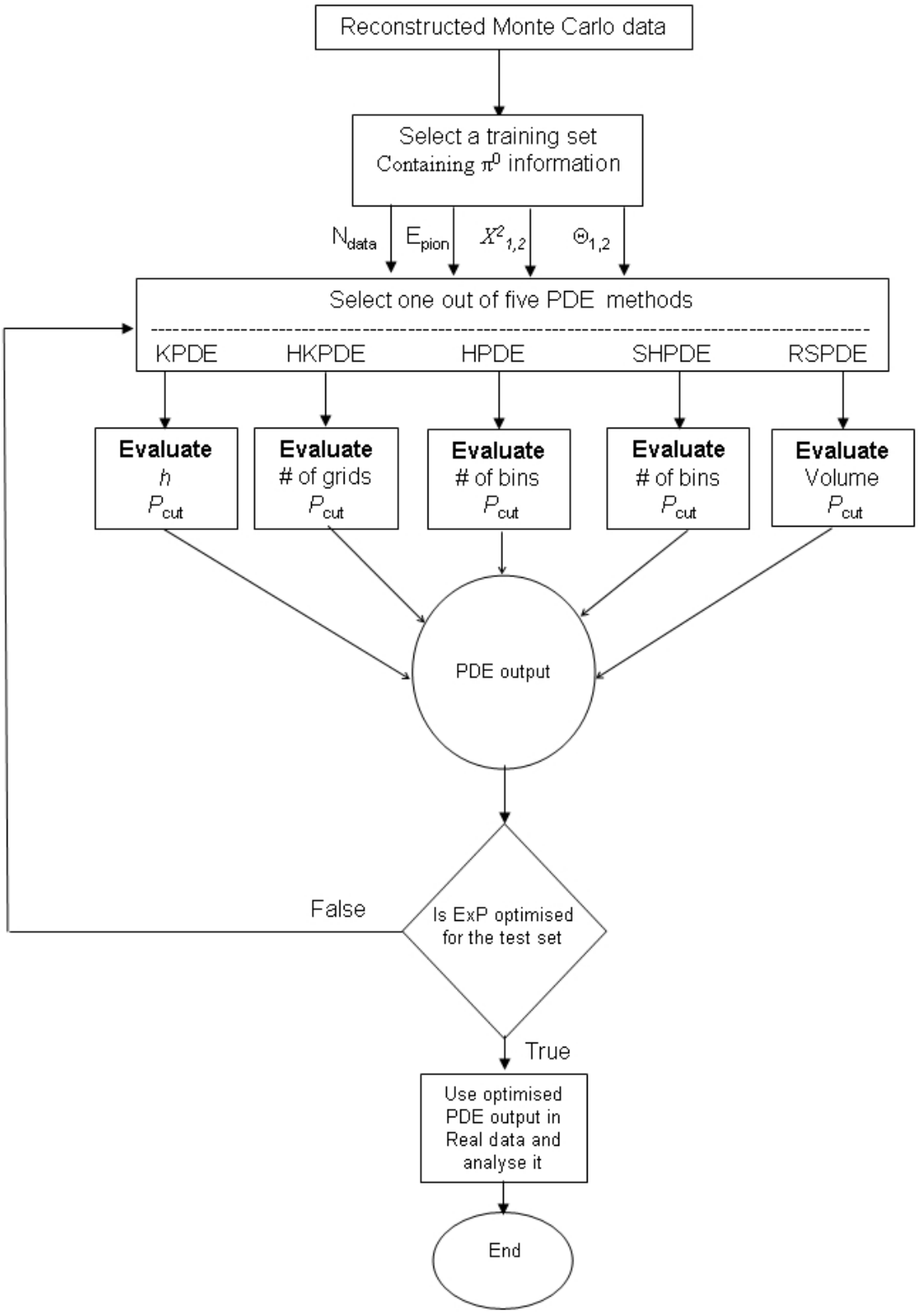

Figure 6.1: Flowchart of the PDE programs

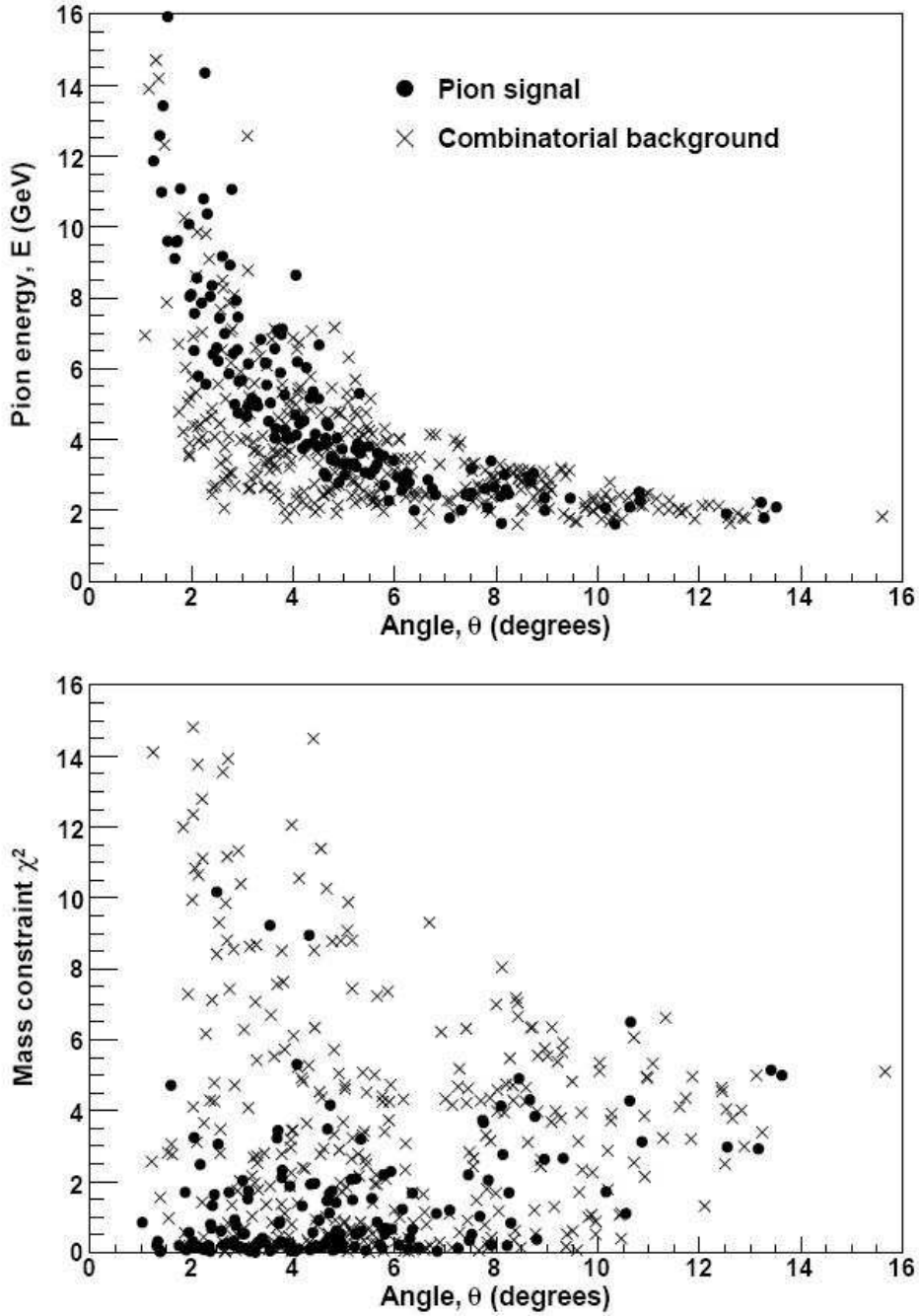

Figure 6.2: Parameters used in the discrimination between signal and background; pion candidates are more likely to be signal if they have smaller values of θ and χ^2, energy provides some discrimination information as the distribution of θ is energy dependent.

6.2.1 $\varepsilon \times \mathcal{P}$ Performance

The mean maximum $\varepsilon \times \mathcal{P}$ values for each PDE method are shown in Figure 6.3 for different number of data. The Figure also includes a comparison with results from a feed-forward artificial neural network (ANN); one hidden layer with seven nodes was found to be optimal for this data. The error bars represent the standard deviation of the twenty mean values. Results for the KPDE, HKPDE, and RSPDE are not computed at high statistics due to the very large run-times necessary for these methods.

From Figure 6.3, it can be seen that given enough statistics all methods perform equally well, raising the $\varepsilon \times \mathcal{P}$ value from an initial 0.32 to about 0.40, this represents a significant improvement in the selected signal significance. At lower statistics, results begin to fluctuate, this is apparent in the larger standard deviations, and scattering of the points.

The form of the results can be explained by considering two main effects. First, the fluctuations in the results, i.e. the standard deviations, appear to be only statistical in nature, i.e. they are independent of the method as each method experiences fluctuations of the same size for the same number of data. Second, some $\varepsilon \times \mathcal{P}$ values exceed 0.4, this does not represent improved performance[1] , but instead over-training during optimisation using the test data set.

At the lowest statistics, $d = 1$ ($n = 50$), the three poorest performers are those methods involving histogramming, though only the simple histogram method has a performance that is significantly lower, this method shows no discrimination performance for this number of data. At first inspection, the highest performer appears to be the KPDE method, but assuming that its high $\varepsilon \times \mathcal{P}$ value of about 0.42 is due to over-training then we can conclude that the performance of this method when applied to real data is actually poorer and a resultant efficiency calculation may contain a significant systematic error.

The results from the neural network appears to exhibit, on average, no over-trainer nor performance loss; the $\varepsilon \times \mathcal{P}$ values are consistent throughout the range of data scale. This is not surprising as neural networks are well known for their good discrimination behavior. However, the 'black box' nature of the neural network is not preferred by researchers due to the difficulty in assessing systematic errors.

[1] We know, from the high statistics data, that the optimal performance is measured as $\varepsilon \times \mathcal{P} = 0.40$.

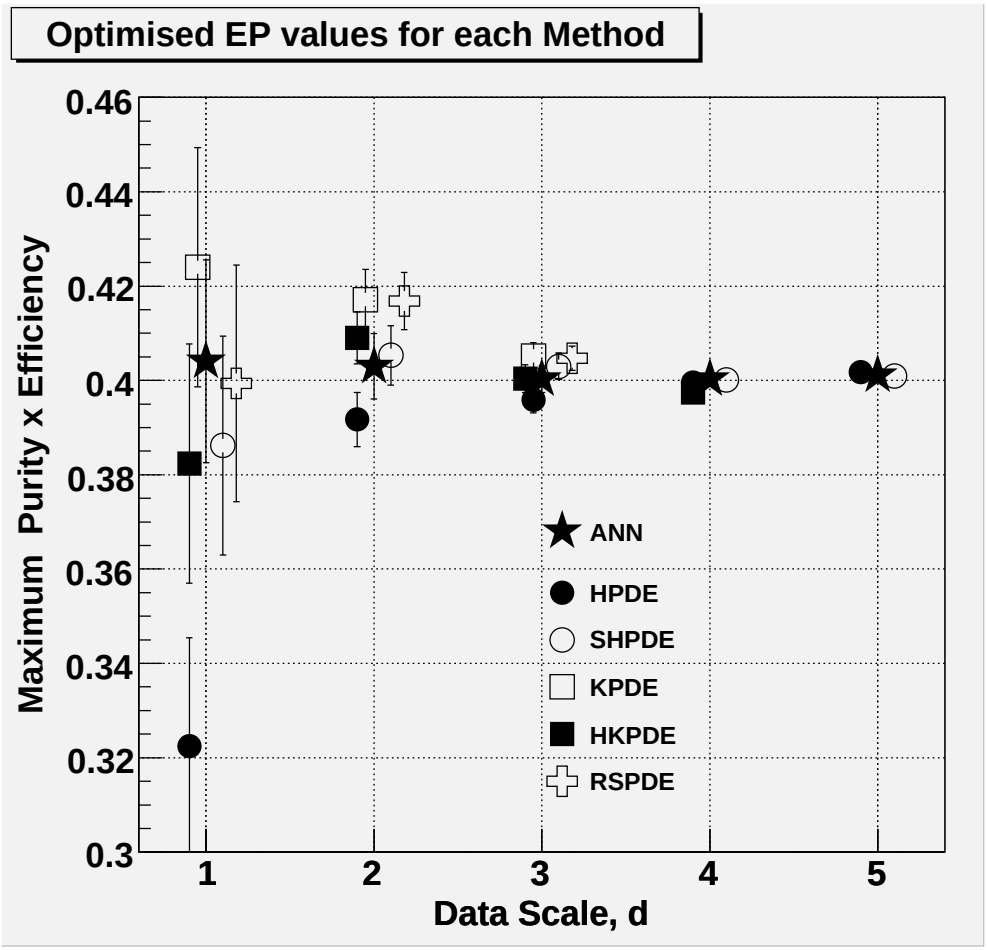

Figure 6.3: Comparison of optimal $\varepsilon \times \mathcal{P}$ values for five PDE methods. The total number of data points scales as $n = 5 \times 10^d$. Results from an artificial neural network (ANN) are also shown.

6.2.2 Timing

Run-time can be an important factor when choosing a PDE algorithm, especially in shared environments where, for example, jobs may be controlled by a queue manager, or on desktop computers that may have low cpu power. The average run-times[2] for each PDE method, including appropriate scales representing the number of repetitions employed in the optimisation procedure, are shown in Figure 6.4 as a function of the data scale. For the Kernel and Range-Search methods the highest data scale values are predictions.

From this log-log plot, the $O(n)$ time complexity of the HPDE, SHPDE, HKPDE, and ANN methods, and the $O(n^2)$ time complexity of the KPDE and RSPDE can be seen. At largest data scale, $d = 5$, ($n = 500{,}000$ pion candidates), the RSPDE and KPDE methods are predicted to take of the order of days, and weeks respectively to train. By far the fastest methods are the simple and smoothed histogrammed PDEs that take only a few minutes to process 500,000 three-dimensional data points. However, at low statistics, the run-times of any method are only a few seconds or less and so run-time for small data sets is not likely to be an issue for the researcher. Also, at low statistics, techniques to speed up $O(n^2)$ algorithms may not be affected, as is the case of the HKPDE method where the procedure of histogramming introduces and extra optimisation parameter which in turns makes the algorithm relatively slower at small and medium data scales.

[2] The CPU times are recorded using an Intel 3.73 GHz Pentium 4 PC running the Linux operating system.

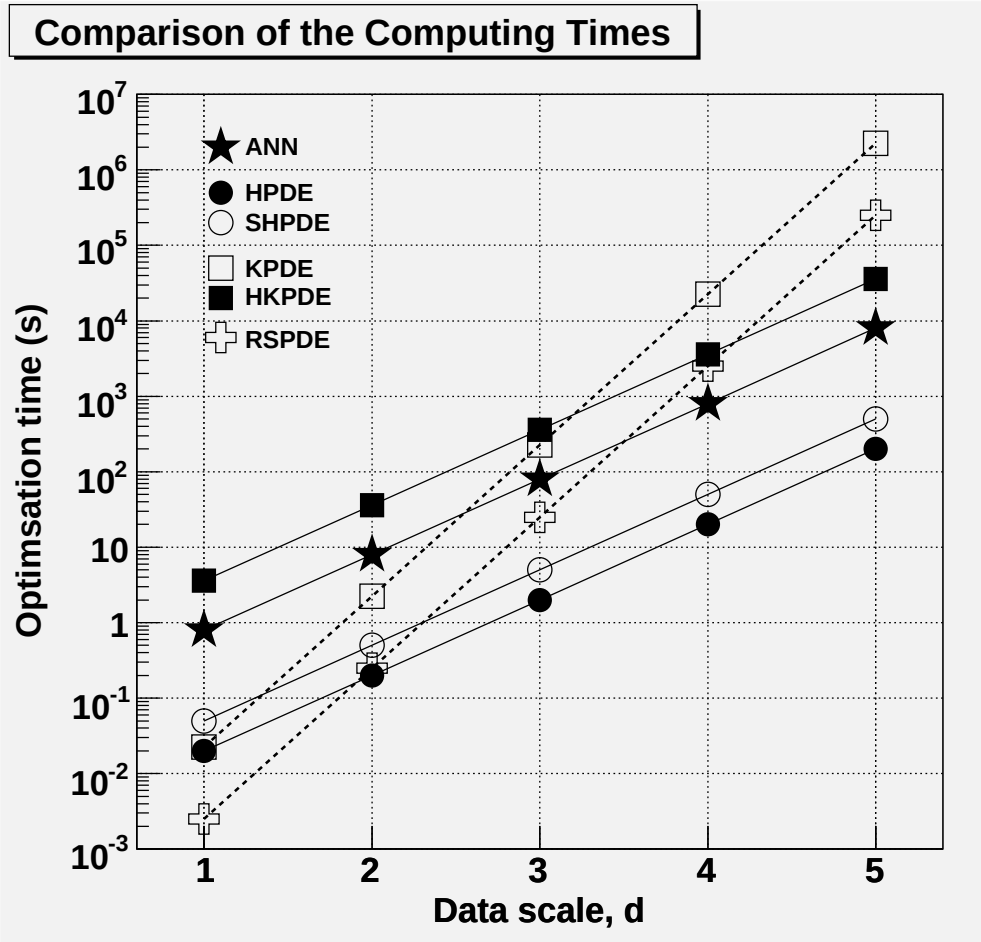

Figure 6.4: Comparison of average run-times (Intel 3.73 GHz Pentium 4) for five PDE methods. The total number of data points scales as $n = 5 \times 10^d$. The time complexity of a method is seen to be either $O(n)$ or $O(n^2)$.

Chapter 7

Summary and Conclusion

It is found, for a real high energy physics discrimination problem, in three-dimensions, that a number of PDE methods perform equally given enough statistics. At very high statistics, above 10,000 events, a simple histogramming method may be sufficient, and is a few orders of magnitude faster than basic implementations of other PDE methods. Histogramming also carries the advantage that the results can be visualised directly.

As statistics reduce, PDE performance becomes unstable sometimes resulting in over-training and sometimes in under-training; in either case the discrimination performance in real data will be diminished, and efficiency calculations will contain errors. This fundamental problem applies equally to all the PDE methods studied in this thesis down to 50 events. The exception is HPDE method that fails to discriminate with this number of events.

At low statistics, the histogram PDE method does not perform well unless additional smoothing is applied. Here, the Kernel and Range-Searching PDE methods show significant improvements. Although advanced implementations of some PDE methods are available, for example in the ROOT package[12], the simplicity and performance of the KPDE and RSPDE methods in their basic forms may be attractive to researchers who prefer, for various reasons, to implement routines themselves. These algorithms are also very fast if the number of data are less than 1,000. Also, as these two algorithms are conceptually quite different with respect to the way they form estimates of feature functions, they can be used as a systematic check on each other.

List of References

[1] Andreas Höcker et. al, TMVA toolkit for parallel multivariate data analysis, http://tmva.sourceforge.net/

[2] D. Buskulic et. al, Performance of the ALEPH detector at LEP, Nucl. Instr. and Meth. A 360 (1995) 481-506.

[3] J. Knobloch and P. Norton, (ALEPH Collab.), Status of Reconstruction Algorithms for ALEPH, Draft ALEPH note.

[4] D. Buskulic et al. (ALEPH collab.), CERN-PPE / 95-089 (1995), Submitted to Z. Phys. C.

[5] CERN Program Library, long writeup W5013.

[6] A. Askew et. al, Event Selection Using Adaptive Gaussian Kernels, PHYSTAT2003, SLAC, Stanford, California, September 8-11, 2003. http://www.slac.stanford.edu/econf/C030908/papers/TUHT002.pdf

[7] K. Cranmer, Kernel estimation in high-energy physics, Computer Physics Communications 136 (2001) 198-207.

[8] L. Holmstrom et. al, A new multivariate technique for top quark search, Computer Physics Communications 88 (1995) 195-210.

[9] T. Carlia, B. Koblitza, A multi-variate discrimination technique based on range-searching, Nucl. Instr. and Meth. A 501 (2003) 576-588.

[10] T. Sjöstrand, Computer Physics Communications 67 (1994) 74.

[11] A. Heister et. al, Inclusive production of the eta and omega mesons in Z decays, and the muonic branching ratio of the omega, Physics Letters B 528 (2002) 19-33.

[12] ROOT, An Object-Oriented Data Analysis Framework http://root.cern.ch/

Appendix A

π^0Reconstruction from Photon Pairs γ_1,γ_2

Consider the decay process of the neutral pion, $\pi^0 \rightarrow \gamma + \gamma$. Invariant mass is defined for two body decays, for this case two γs, as:

$$\begin{aligned} M^2 &= E^2 - P^2 \\ &= (E_1 + E_2)^2 - (P_{x_1} + P_{x_2})^2 - (P_{y_1} + P_{y_2})^2 - (P_{z_1} + P_{z_2})^2 \end{aligned} \quad \text{(A.1)}$$

where E_1 and E_2 are the energies of photons that can be found by:

$$E_i^2 = P_i^2 + M_\gamma^2 \quad \text{(A.2)}$$

here $i = 1, 2$ and $P_i^2 = P_{x_i}^2 + P_{y_i}^2 + P_{z_i}^2$. $P_{x,y,z}$ is the momentum components of sum of the photons. M_γ is assumed to be the mass of the photon, thus $M_\gamma = 0$. For π^0 we can perform reconstruction from two photons. From the conservation of momentum:

$$\vec{P}_{\pi^0} = \vec{P} = \vec{P}_1 + \vec{P}_2 \quad \text{(A.3)}$$

Now we can find P^2 by the Equation A.3 and definition of dot product of two vectors:

$$\vec{P}.\vec{P} = P^2 = P_1^2 + P_2^2 + 2P_1P_2\cos\theta_{12} \quad \text{(A.4)}$$

where θ_{12} is the angle between two photons. γ_1 and γ_2 are assigned to zero mass, so $E = P$.

Thus Equation A.4 becomes:

$$P^2 = E_1^2 + E_2^2 + 2E_1E_2\cos\theta_{12} \quad \text{(A.5)}$$

If we substitute the final result of P^2 into Equation A.1, then:

$$M^2 = 2E_1E_2(1 - \cos\theta_{12}) \quad \text{(A.6)}$$

We obtain measurements of E_1, E_2 and θ_{12} from the electromagnetic calorimeter. So invariant mass of two photon pairs can be set up by Equation A.6

Appendix B

χ^2 and Lagrange Multiplicator

The calculation is done with the well known technique of Lagrange Multiplicator. Let ω_1 and ω_2 be measured energies of two photons and θ_{12} be opening angle, the reconstructed squared invariant mass of the $\pi^0$is:

$$\mu^2 = 2\omega_1\omega_2(1 - cos\theta_{12}) \tag{B.1}$$

Now we can look at for E_1 and E_2, the best evaluation of energies of the two gammas, minimising the chi-square form:

$$\chi^2 = [(E_1 - \omega_1)/\sigma_1]^2 + [(E_2 - \omega_1)/\sigma_2]2 + \lambda[m^2 - 2E_1E_2(1 - cos\theta_{12})] \tag{B.2}$$

Where σ_1 and σ_2 are errors on photon energy, m is true π^0mass and λ is the Lagrange multiplicator.

B.1 Method of Lagrange Multipliers for Maxima and Minima

A method for obtaining the relative maximum or minimum values of a function $F(x, y)$ subjected to a constraint condition $H(x, y) = 0$, consists of the formation of the auxilary function:

$$G(x, y) = F(x, y) + \lambda H(x, y) \tag{B.3}$$

subject to the conditions:

$$dG/dx = 0 \qquad dG/dy = 0 \tag{B.4}$$

which is necessary conditions for relative maximum or minimum. The parameter λ, which is independent of x, y, is called Lagrange multiplier.The method can be generalized for more than two variables and one constraint functions.